Chapter 1

Introduction

The discussion in this first chapter will give us a common reference to present the results on the intersection of probabilistic methods and graph searching games. As the name suggests, this is a book on graphs and probability (we will deal with the searching part more explicitly in the following chapters). With a combinatorial audience in mind, we devote a few brief pages to summarize some notation from graph theory, and spend more time covering a few elementary but key theorems in discrete probability. An advanced reader may safely skim these pages and move directly to Chapter 2; this chapter may be used, nevertheless, as a quick reference for statements of key facts (like the Chernoff bounds) that we freely use later. More involved tools, such as martingales or the differential equations method, will be introduced in later chapters as needed.

Some basic notation comes first. The set of *natural numbers* (excluding 0 for notation simplicity, although this notation often includes 0) is written $\mathbb{N}$ while the *rationals* and *reals* are denoted by $\mathbb{Q}$ and $\mathbb{R}$, respectively. If n is a natural number, then define

$$[n] = \{1, 2, \ldots n\}.$$

The *Cartesian product* of two sets A and B is written $A \times B$. The difference of two sets A and B is written $A \setminus B$. We use the notation $\log n$ for the logarithm in the natural base.

1.1 Graphs

Graphs are our main objects of study. For further background in graph theory, the reader is directed to any of the texts [34, 76, 180].

A *graph* $G = (V, E)$ is a pair consisting of a *vertex set* $V = V(G)$, an *edge set* $E = E(G)$ consisting of pairs of vertices. Note that E is taken as a multiset, as its elements may occur more than once. We write uv if u and v form an edge, and say that u and v are *adjacent* or *joined*. For consistency, we will use the former term only. We refer to u and v as *endpoints* of the edge uv. The *order* of a graph is $|V(G)|$, and its *size* is $|E(G)|$. Graphs are often depicted by their drawings; see Figure 1.2.

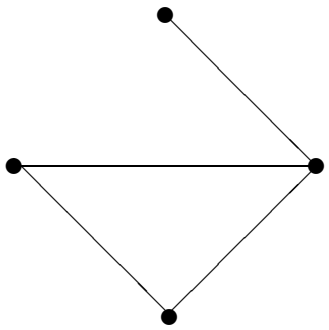

FIGURE 1.1: An example of a graph of order and size 4.

A *loop* is an edge whose endpoints are equal. *Multiple edges* are edges having the same pair of endpoints. If u and v are the endpoints of an edge, then we say that they are *neighbors*. The *neighborhood* $N(v) = N_G(v)$ of a vertex v is the set of all neighbors of v. We usually restrict our attention to *simple graphs*; that is, graphs without loops and multiple edges. Further, we only consider finite graphs.

The *degree* of a vertex v in G, written $\deg_G(v)$, is the number of neighbors of v in G; that is, $\deg_G(v) = |N(v)|$. We will drop the subscript G if the graph is clear from context. The number $\delta(G) = \min_{v \in V(G)} \deg(v)$ is the *minimum degree* of G, and the number $\Delta(G) = \max_{v \in V(G)} \deg(v)$ is the *maximum degree* of G. A graph is *k-regular* if each vertex has degree k.

The *complement* $\overline{G}$ of a graph G is the graph with vertex set $V(\overline{G}) = V(G)$ and edge set $E(\overline{G})$ defined by $uv \in E(\overline{G})$ if and only if $uv \notin E(G)$. See Figure 1.2. A *clique* (sometimes called a *complete graph*) is a set of pairwise-adjacent vertices. The clique of order n is denoted by K_n. An *inde-*

pendent set (sometimes called an *empty graph*) is a set of pairwise-nonadjacent vertices. Note that an independent set is the complement of a clique.

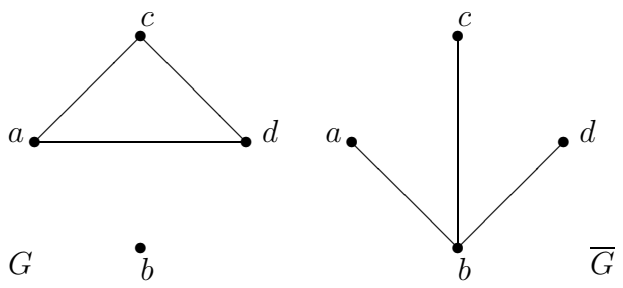

FIGURE 1.2: The graph G and its complement $\overline{G}$.

A graph G is *bipartite* if $V(G) = X \cup Y$, where $X \cap Y = \emptyset$, and every edge is of the form xy, where $x \in X$ and $y \in Y$; here X and Y are called *partite sets*. The *complete bipartite graph* $K_{m,n}$ is the graph with partite sets X, Y with $|X| = m$, $|Y| = n$, and edge set

$$E = \{xy : x \in X, y \in Y\}.$$

A graph $G' = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. We say that G' is a *spanning subgraph* if $V' = V$. If $V' \subseteq V$, then

$$G[V'] = (V', \{uv \in E : u, v \in V'\})$$

is the subgraph of G *induced* by V'. Similarly, if $E' \subseteq E$, then $G[E'] = (V', E')$ where

$$V' = \{v \in V : \text{there exists } e \in E' \text{ such that } v \in e\}$$

is an *induced subgraph of* G *by* E'. Given a graph $G = (V, E)$ and a vertex $v \in V$, we define $G - v = G[V \setminus \{v\}]$. For an edge e, $G - e$ is the subgraph formed by deleting e.

An *isomorphism* from a graph G to a graph H is a bijection $f : V(G) \to V(H)$ such that $uv \in E(G)$ if and only if $f(u)f(v) \in E(H)$. G is *isomorphic* to H, written $G \cong H$, if there is an isomorphism from G to H. See Figure 1.3 for two isomorphic graphs.

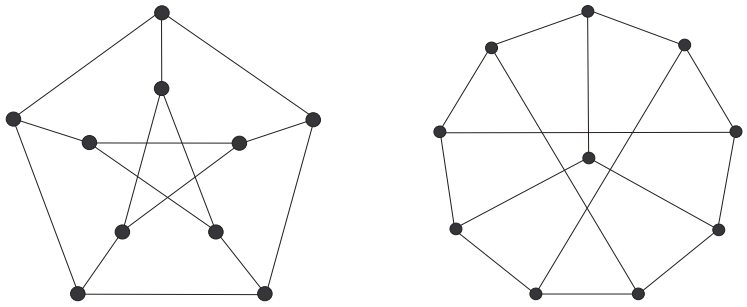

FIGURE 1.3: Two graphs isomorphic to the Petersen graph.

A *walk* in a graph $G = (V, E)$ from vertex u to vertex v is a sequence $W = (u = v_0, v_1, \ldots, v_l = v)$ if $v_i v_{i+1} \in E$ for $0 \leq i < l$. The *length* $l(W)$ of a walk W is the number of vertices in W *minus 1* (that is, the number of edges). A walk is *closed* if $v_0 = v_l$. A *path* is a walk in which the internal vertices are distinct. The path of order n is denoted by P_n. A *cycle* is a closed path of length at least 3. We use the notation C_n for a cycle of order n. A graph G is *connected* if there is a walk (equivalently, a path) between every pair of vertices; otherwise, G is *disconnected*. See Figure 1.4. A *connected component* (or just *component*) of a graph G is a maximal connected subgraph. A connected component consisting of a single vertex is called an *isolated vertex*. A vertex adjacent to all other vertices is called *universal*.

A *forest* is a graph with no cycle. A *tree* is a connected forest; hence, every component of a forest is a tree. Each tree on n vertices has size $n - 1$. An *end-vertex* is a vertex of degree 1; note that every *nontrivial* tree (that is, a tree of order at least 2) has at least two end-vertices. A *spanning tree* is a spanning subgraph that is a tree. The graph P_n and an n-vertex *star* $K_{1,n-1}$ are trees. A *hypercube* of dimension n, written Q_n, has vertices elements of $\{0, 1\}^n$, with two vertices adjacent if they differ in exactly one coordinate. In particular, Q_n has order 2^n and size $n2^{n-1}$.

For distinct vertices u and v, the *distance* between u and v, written $d_G(u, v)$ (or just $d(u, v)$) is the length of a shortest path connecting u and v if such a path exists, and ∞, otherwise. We take the distance between a vertex and itself to be 0. The *diameter* of a connected graph G, written $\mathrm{diam}(G)$, is the maximum of all distances between vertices. If the graph

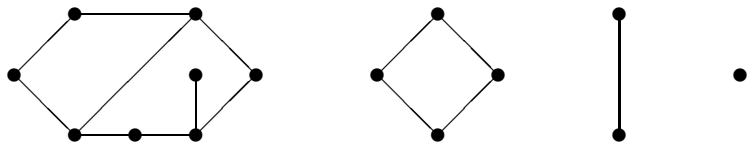

FIGURE 1.4: A disconnected graph with 4 components.

is disconnected, then diam(G) is ∞. For a nonnegative integer r and vertex u in G, define $N_r(u)$ to be set of those vertices of distance r from u in G. Note that $N_0(u) = \{u\}$ and $N_1(u) = N(u)$.

The *breadth-first search* (BFS) process is a graph search algorithm that begins at the root vertex v and explores all the neighboring vertices. Then for each of those neighboring vertices, it explores their unexplored neighbors, and so on, until it explores the whole connected component containing vertex v. Formally, the algorithm starts by putting vertex v into a *FIFO queue*; that is, *First In, First Out*. In each round, one vertex is taken from the queue and all neighbors that have not yet been discovered are added to the queue. The process continues until the queue is empty. It may be shown that the BFS process naturally yields the breadth-first search tree.

A *dominating set* of a graph $G = (V, E)$ is a set $U \subseteq V$ such that every vertex $v \in V \backslash U$ has at least one neighbor in U. The *domination number* of G, written $\gamma(G)$, is the minimum cardinality of a dominating set in G. Note that the vertex set V is a dominating set. However, it is usually possible to find a much smaller dominating set (for example, consider a graph with a universal vertex).

A *matching* in a graph G is a 1-regular subgraph. A matching is *maximal* if it cannot be extended by adding an edge. A matching is *maximum* if it contains the largest possible number of edges. A *perfect matching* in a graph G is a matching in G that is a spanning subgraph of G.

The *line graph* of a graph G, written $L(G)$, is the graph whose vertices are the edges of G, with $ef \in E(L(G))$ when $e = uv$ and $f = vw$ are both in $E(G)$. See Figure 1.5 for an example. For graphs G and H, define the *Cartesian product* of G and H, written $G \square H$, to have vertices $V(G) \times V(H)$, and

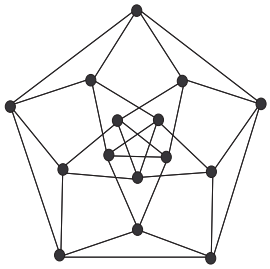

FIGURE 1.5: The line graph of the Petersen graph (see Figure 1.3).

vertices (a, b) and (c, d) are adjacent if $a = c$ and $bd \in E(H)$ or $ac \in \mathbb{E}(G)$ and $b = d$.

We can assign a direction to each edge of a graph G. A simple *directed graph* (or *digraph*) $G = (V, E)$ is a pair consisting of a vertex set $V = V(G)$ and an edge set $E = E(G) \subseteq \{(x, y) : x, y \in V(G), x \neq y\}$. See Figure 1.6; we use the arrow notation to depict an edge pointing from vertex to vertex. The *in-degree* of a vertex v, written $\deg^-(v)$, is the number

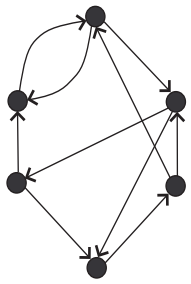

FIGURE 1.6: An example of a digraph.

of in-neighbors of v; that is,

$$\deg^-(v) = |\{u \in V : (u, v) \in E\}|.$$

Similarly, the *out-degree* of a vertex v, written $\deg^+(v)$, is the number of out-neighbors of v; that is,

$$\deg^+(v) = |\{u \in V : (v, u) \in E\}|.$$

1.2 Probability

We next introduce some basic definitions and theorems from discrete probability theory. For more details and any proofs not provided here, see, for example, [8, 104].

Definitions

The set of possible outcomes of an experiment is called the *sample space* and is denoted by Ω. An *elementary event* is an event that contains only a single outcome in the sample space. For example, a coin is tossed. There are two possible outcomes: heads (H) and tails (T), so $\Omega = \{H, T\}$. We might be interested in the following events:

(i) the outcome is H,

(ii) the outcome is H or T,

(iii) the outcome is not H, and so on.

Note that we think of *events* as subsets of Ω.

A collection $\mathcal{F}$ of subsets of Ω is called a *σ-field* if it satisfies the following conditions:

(i) $\emptyset \in \mathcal{F}$,

(ii) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$, and

(iii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

The smallest σ-field associated with Ω is the collection $\mathcal{F} = \{\emptyset, \Omega\}$. If A is any subset of Ω, then $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ is a σ-field. The *power set* of Ω, which contains all subsets of Ω, is obviously a σ-field. For reasons beyond the scope of this book, when Ω is infinite, its power set is too large for probabilities to be assigned reasonably to all its members. Fortunately, we are going to deal with finite graphs and related structures only, and so from now on it will always be assumed that a σ-field is the power set of Ω.

A *probability measure* $\mathbb{P}$ on $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-field, is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ such that

(i) $\mathbb{P}(\Omega) = 1$, and

(ii) if $A_1, A_2, \ldots$ is a sequence of pairwise disjoint events, then

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

Basic Properties

Below we present a few elementary properties of a probability space.

Theorem 1.2.1. *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then for any $A, B \in \mathcal{F}$ we have that*

(i) $\mathbb{P}(\emptyset) = 0$,

(ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

(iii) if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$, and

(iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

The last equality can be generalized to the *Inclusion-Exclusion Principle*. Let $A_1, A_2, \ldots, A_n$ be events, where $n \geq 2$.

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\ &\quad + \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).\end{aligned}$$

If $\mathbb{P}(B) > 0$, then the *conditional probability* that A occurs given that B occurs is defined to be

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

We state a number of facts about conditional probabilities that will be used (sometimes implicitly) in later discussions.

Theorem 1.2.2 (Law of total probabilities)**.** *If $A \in \mathcal{F}$ is an event with $P(A) > 0$, and $\{A_i\}_{i=1}^n$ is a partition of A, then we have that*

$$\mathbb{P}(B|A) = \frac{1}{\mathbb{P}(A)} \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

In particular, taking $A = \Omega$, we derive the following corollary.

Corollary 1.2.3. *If $\{A_i\}_{i=1}^n$ is a partition of the sample space Ω, then we have that*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

We also have the following.

Theorem 1.2.4 (Chain law)**.** *If $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then for any event $B \in \mathcal{F}$ such that $\mathbb{P}\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cap B\right) > 0$, we have that*

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i | B\right) &= \mathbb{P}(A_1|B) \cdot \mathbb{P}(A_2|A_1 \cap B) \\ &\quad \cdots \mathbb{P}\left(A_n \Big| \left(\bigcap_{i=1}^{n-1} A_i\right) \cap B\right). \end{aligned}$$

In particular, taking $B = \Omega$, we obtain the following corollary.

Corollary 1.2.5 (Principle of deferred decision)**.** *If we have $A_1, A_2, \ldots, A_n \in \mathcal{F}$ such that $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$, then*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \cdots \mathbb{P}\left(A_n \Big| \bigcap_{i=1}^{n-1} A_i\right).$$

We will make use of the following equalities.

Theorem 1.2.6 (Bayes' law)**.** *If $\{A_i\}_{i=1}^n$ is a partition of the sample space Ω and $B \in \mathcal{F}$, then for any $j \in [n]$*

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Proof. Note that

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \quad \square$$

The following elementary fact, also known as the *union bound*, proves useful.

Lemma 1.2.7 (Boole's inequality)**.** *If $A_1, A_2, \ldots, A_n$ are events, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

We also mention a generalization of Boole's inequality, known as Bonferroni's inequality (see Lemma 1.2.8). This inequality can be used to find upper and lower bounds on the probability of a finite union of events. Boole's inequality is recovered by setting $k = 1$. If $k = n$, then equality holds, and the resulting identity is the inclusion-exclusion principle.

Lemma 1.2.8 (Bonferroni inequalities)**.** *Let*

$$B_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

for all integers $k \in [n]$. Then for odd $k \in [n]$ we have that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{j=1}^{k} (-1)^{j-1} B_j,$$

and for even $k \in [n]$ we have that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{j=1}^{k} (-1)^{j-1} B_j.$$

Useful Distributions

Here are some discrete probability distributions that we use throughout.

(i) *Bernoulli*(p): Fix $p \in (0, 1)$.

$$\mathbb{P}(X = x) = \begin{cases} p, & \text{if } x = 1; \\ 1 - p, & \text{if } x = 0. \end{cases}$$

Here p is the *success probability* and $1 - p$ is the *failure probability*.

(ii) *Binomial*(n, p) (we will use $\mathrm{Bin}(n, p)$ instead): Fix $p \in (0, 1)$ and $n \in \mathbb{N}$. Let X be the number of successes in n independent repetitions of the same Bernoulli(p) trial. Then we have that

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad 0 \leq k \leq n.$$

(iii) *Hypergeometric*(N, K, n): Let $N, K, n, k \in \mathbb{N}$. The hypergeometric distribution describes the probability of k successes in n draws, without replacement, from a population of size N that contains exactly K successes, wherein each draw is either a success or a failure. A random variable X follows the hypergeometric distribution if

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

(iv) *Geometric*(p): Fix $p \in (0, 1)$. Let X be the *waiting time* (that is, the number of trials) for the first success in independent repetitions of the same Bernoulli(p) trial. Then it follows that

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \qquad k = 1, 2, \ldots.$$

Note that $\mathbb{P}(X > k) = (1-p)^k$. Note also that the geometric distribution is "memory-less"; that is, for $r > k$,

$$\begin{aligned} \mathbb{P}(X > r | X > k) &= \frac{\mathbb{P}(X > r)}{\mathbb{P}(X > k)} = \frac{(1-p)^r}{(1-p)^k} \\ &= (1-p)^{r-k} = \mathbb{P}(X > r-k). \end{aligned}$$

(v) *Poisson*(λ) (we will use Po(λ) instead): Fix $\lambda > 0$.

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k = 0, 1, 2, \ldots.$$

1.3 Asymptotic Notation and Useful Inequalities

Since many of our results will be asymptotic, in this section we recall some asymptotic notation and some inequalities that we frequently use.

Asymptotic Notation

Let $f(n)$ and $g(n)$ be two functions whose domain is some fixed subset of $\mathbb{R}$, and assume that $g(n) > 0$ for all n. We say that f *is of order at most* g, written $f(n) = O(g(n))$, if there exist constants $A > 0$ and $N > 0$ such that for all $n > N$, we have that

$$|f(n)| \leq A|g(n)|.$$

Observe that $f(n)$ could be negative or even oscillate between negative and positive values (for example, $2\sin(3n)n^2 = O(n^2)$). We say that f *is of order at least* g, written $f(n) = \Omega(g(n))$, if there exist constants $A > 0$ and $N > 0$ such that for all $n > N$,

$$f(n) \geq Ag(n).$$

Finally, we say that f *is of order* g, written $f(n) = \Theta(g(n))$, if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

Note that for a polynomial $p(n)$ of degree k, $p(n) = \Theta(n^k)$. Here are some useful properties of the O- and Ω- notation, which are by no means exhaustive.

Theorem 1.3.1. *For positive functions $f(n)$ and $g(n)$ we have the following.*

(i) $O(f(n)) + O(g(n)) = O(f(n) + g(n))$.

(ii) $f(n)O(g(n)) = O(f(n)g(n))$ *and* $f(n)\Omega(g(n)) = \Omega(f(n)g(n))$.

(iii) If $f(n) = O(1)$, *then* $f(n)$ *is bounded by a constant.*

(iv) $n^r = O(n^s)$ *for any real numbers* r, s *with* $r \leq s$.

(v) $n^r = O(a^n)$ *for any real numbers* r, a *with* $a > 1$.

(vi) $\log n = O(n^r)$ *for any real number* $r > 0$ *(note that this could be the logarithm to any base, since* $\log_a x = \frac{\log_b x}{\log_b a} = \Theta(\log_b x)$ *for any* $a, b > 1$*).*

(vii) $\log\log n = O(\log n)$ *and* $\log\log\log n = O(\log\log n)$.

We say that f is *of order smaller than* g, written $f(n) = o(g(n))$ or $f(n) \ll g(n)$, if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Note that we do not control the sign of function $f(n)$ while in the next definition we do. The function f is *of order larger than* g, written $f(n) = \omega(g(n))$ or $f(n) \gg g(n)$, if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty.$$

Finally, f is *asymptotically equal* to g, written $f(n) \sim g(n)$ or $f(n) = (1 + o(1))g(n)$, if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1.$$

For more details about asymptotic notation see, for example, [177].

Useful Inequalities

We collect a few inequalities that will be used throughout. For all $x \in \mathbb{R}$, $e^x \geq 1+x$. Hence, for all $x \in \mathbb{R}$ that are positive, we have that $\log(1 + x) \leq x$. For the factorial function $n!$, we have

$$\left(\frac{n}{e}\right)^n \leq n! \leq en\left(\frac{n}{e}\right)^n.$$

Stirling's formula, as stated in the following lemma, is often useful for our estimates.

Lemma 1.3.2.

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

More precisely,

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\lambda_n}, \text{ with } \frac{1}{12n+1} < \lambda_n < \frac{1}{12n},$$

so

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + O(n^{-3})\right).$$

For the binomial coefficient $\binom{n}{k}$, we have the inequalities

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

For the middle binomial coefficient $\binom{2m}{m}$, we have the better estimate

$$\frac{2^{2m}}{2\sqrt{m}} \leq \binom{2m}{m} \leq \frac{2^{2m}}{\sqrt{2m}},$$

and from Stirling's formula the asymptotic behavior may be obtained as follows:

$$\binom{2m}{m} = \frac{(2m)!}{(m!)^2} \sim \frac{\sqrt{4\pi m}(2m/e)^{2m}}{\left(\sqrt{2\pi m}(m/e)^m\right)^2} = \frac{2^{2m}}{\sqrt{\pi m}}.$$

1.4 Random Graphs

In this section, we supply five examples of probability spaces. These notions of *random graphs*, especially the second and to a lesser extent the third and fourth, will be central to our discussion.

Let Ω be the family of all graphs with n vertices and exactly M edges, $0 \leq M \leq \binom{n}{2}$. To every graph $G \in \Omega$ we assign a uniform probability; that is,

$$\mathbb{P}(\{G\}) = \binom{\binom{n}{2}}{M}^{-1}.$$

We denote this associated probability space by $\mathbb{G}(n, M)$.

Now, let $0 \leq p \leq 1$ and let Ω be the family of all graphs on n vertices. To every graph $G \in \Omega$ we assign a probability

$$\mathbb{P}(\{G\}) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}.$$

Note that this indeed is a probability measure, since

$$\sum_{G} \mathbb{P}(\{G\}) = \sum_{m=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} = (p+(1-p))^{\binom{n}{2}} = 1.$$

We denote this probability space by $\mathbb{G}(n, p)$. The space $\mathbb{G}(n, p)$ is often referred to as the *binomial random graph* or *Erdős-Rényi random graph*. Note also that this probability space can informally be viewed as a result of $\binom{n}{2}$ independent coin flips, one for each pair of vertices u, v, where the probability of success (that is, adding an edge uv) is equal to p. Let us also note that if $p = 1/2$, then $\mathbb{P}(\{G\}) = 2^{-\binom{n}{2}}$ for any graph G on n vertices. We obtain a uniform probability space.

Next, let Ω be the family of all d-regular graphs on n vertices, where $0 \leq d \leq n-1$ and dn is even. (Note that the condition dn is needed; otherwise, $\Omega = \emptyset$.) To every graph $G \in \Omega$ we assign a uniform probability; that is,

$$\mathbb{P}(\{G\}) = \frac{1}{|\Omega|}.$$

We refer to this space as the *random regular graph of degree* d, and write $\mathbb{G}_{n,d}$.

As typical in random graph theory, we shall consider only asymptotic properties of $\mathbb{G}(n, M)$ and $\mathbb{G}(n, p)$ as $n \to \infty$, where $M = M(n)$ and, respectively, $p = p(n)$ may and usually do depend on n. For $\mathbb{G}_{n,d}$ we typically concentrate on d being a constant but it is also interesting to consider $d = d(n)$ tending to infinity with n. We say that an event in a probability space holds *asymptotically almost surely* (*a.a.s.*) if its probability tends to one as n goes to infinity. For mode details see, for example, the two classic books [35, 115] or more recent monograph [94]. Random d-regular graphs are also discussed in the survey [184].

Finally, we introduce the *random geometric graph* $\mathbb{G}(\mathcal{X}_n, r_n)$, where (i) $\mathcal{X}_n$ is a set of n points located independently uniformly at random in $[0, \sqrt{n}]^2$, (ii) $(r_n)_{n\geq 1}$ is a sequence of positive real integers, and (iii) for $\mathcal{X} \subseteq \mathbb{R}^2$ and $r > 0$, the graph $\mathbb{G}(\mathcal{X}, r)$ is defined to have vertex set $\mathcal{X}$, with two vertices connected by an edge if and only if their spatial locations are at Euclidean distance at most r from each other. As before, we shall consider only asymptotic properties of $\mathbb{G}(\mathcal{X}_n, r_n)$ as $n \to \infty$. We will therefore write $r = r_n$, we will identify vertices with their spatial locations, and we will define $\mathbb{G}(n, r)$ as the graph with vertex set $[n]$ corresponding to n locations chosen independently uniformly at random in $[0, \sqrt{n}]^2$ and a pair of vertices within Euclidean distance r appear as an edge.

We will also consider the *percolated random geometric graph* $\mathbb{G}(n, r, p)$, which is defined as a random graph with vertex set $[n]$ corresponding to n locations chosen independently uniformly at random in $[0, \sqrt{n}]^2$, and for each pair of vertices within Euclidean distance at most r we flip a biased coin with success probability p to determine whether there is an edge (independently for each such a pair, and pairs at distance bigger than r never share an edge). In particular, for $p = 1$ we simply have the random geometric graph $\mathbb{G}(n, r)$. Percolated random geometric graphs were recently studied by Penrose [154] under the name *soft random geometric graphs*.

Alternatively, we can scale the space and define the model in $[0, 1]^2$. Of course, results from one model can be translated to the other one and we will make sure that it is always clear which model we have in mind. For more details, see, for example, the monograph [155].

We note that we use the notations $\mathbb{G}(n,p)$, $\mathbb{G}(n,M)$, and $\mathbb{G}(n,r)$, but this will not cause confusion as it will be clear from the context which of the corresponding three models we are discussing.

1.5 Tools: First and Second Moment Methods

Events A, B are *independent* if

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

In general, events $A_1, A_2, \ldots, A_n$ are *independent* if for any $I \subseteq [n]$,

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i).$$

Intuitively, the property of independence means that the knowledge of whether some of the events $A_1, A_2, \ldots, A_n$ occurred does not affect the probability that the remaining events occur.

A *random variable* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable; that is, for any $x \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

Random variables X, Y are *independent* if for every pair of events $\{X \in A\}$, $\{Y \in B\}$, where $A, B \subseteq \mathbb{R}$, we have that

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(\{X \in A\})\mathbb{P}(\{Y \in B\}).$$

Thus, two random variables are independent if and only if the events related to those random variables are independent events.

The First Moment Method

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a finite probability space. The *expectation* of a random variable X is defined as

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

A simple but useful property of expectation is the following.

Lemma 1.5.1 (Linearity of expectation)**.** *For any two random variables X, Y and $a, b \in \mathbb{R}$, we have that*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Linearity of expectation implies that for all random variables $X_1, X_2, \ldots, X_n$ and all $c_1, c_2, \ldots, c_n \in \mathbb{R}$,

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i \mathbb{E}[X_i].$$

This is a simple observation, but a powerful one. It is important to point out that it holds for both dependent and independent random variables.

For an event $A \in \mathcal{F}$, we define the *indicator random variable* as follows:

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

It is evident that

$$\mathbb{E}[I_A] = \sum_{\omega \in \Omega} \mathbb{P}(\omega) I_A(\omega) = \sum_{\omega \in A} \mathbb{P}(\omega) = \mathbb{P}(A).$$

It is common that a random variable can be expressed as a sum of indicators. In such case, the expected value can also be expressed as a sum of expectations of corresponding indicators.

The first moment method that we are about to introduce is a standard tool used in investigating random graphs and the probabilistic method. It is a useful tool to bound the probability that a random variable X satisfies $X \geq 1$.

Theorem 1.5.2 (Markov's inequality)**.** *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and X is a nonnegative random variable, then for all $\varepsilon > 0$,*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Proof. Note that

$$X = XI_{X\geq\varepsilon} + XI_{X<\varepsilon} \geq XI_{X\geq\varepsilon} \geq \varepsilon I_{X\geq\varepsilon}\,.$$

Hence, by linearity of expectation, we have that

$$\mathbb{E}[X] \geq \varepsilon\mathbb{E}[I_{X\geq\varepsilon}] = \varepsilon\mathbb{P}(X \geq \varepsilon),$$

and the theorem holds. □

Markov's inequality has a simple corollary, proved by setting $\varepsilon = 1$.

Corollary 1.5.3 (The first moment method)**.** *If X is a non-negative integer-valued random variable, then*

$$\mathbb{P}(X > 0) \leq \mathbb{E}[X]\,.$$

The Second Moment Method

Next, we will use the variance to bound the probability that a random variable X satisfies $X = 0$. The second moment method that we are about to introduce is another standard tool used in investigating random graphs. Let X be a random variable. The *variance* of X is defined as

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

that is,

$$\mathbb{V}\mathrm{ar}[X] = \sum_{i=1}^{\infty}(x_i - \mathbb{E}[X])^2\mathbb{P}(X = x_i)\,.$$

We collect some elementary properties of the variance. For random variables X and Y, the *covariance* $\mathbb{C}\mathrm{ov}(X, Y)$ is defined as

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

Theorem 1.5.4. *Let X and Y be random variables, and let a and b be real numbers. The variance operator has the following properties.*

(i) $\mathbb{V}\mathrm{ar}[X] \geq 0$.

(ii) $\mathbb{V}\mathrm{ar}[aX + b] = a^2\mathbb{V}\mathrm{ar}[X]$.

(iii)

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X + Y] &= \mathbb{C}\mathrm{ov}(X, X) + \mathbb{C}\mathrm{ov}(Y, Y) + 2\mathbb{C}\mathrm{ov}(X, Y)\\ &= \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y] + 2\mathbb{C}\mathrm{ov}(X, Y).\end{aligned}$$

(iv) If X, Y are independent, then

$$\mathbb{V}\mathrm{ar}[X+Y] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y]\,.$$

The following theorem is a key tool to achieve our goal.

Theorem 1.5.5 (Chebyshev's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If X is a random variable, then for any $\varepsilon > 0$,*

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geq \varepsilon\Big) \leq \frac{\mathbb{V}\mathrm{ar}[X]}{\varepsilon^2}\,.$$

Proof. By Markov's inequality

$$\mathbb{P}(|Y| \geq \varepsilon) = \mathbb{P}(Y^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[Y^2]}{\varepsilon^2}\,.$$

Setting $Y = X - \mathbb{E}[X]$ we derive the desired assertion. □

In combinatorial applications of probability, the following consequence of Chebyshev's inequality plays an important role, as we will develop in later chapters.

Theorem 1.5.6 (The second moment method)**.** *If X is a nonnegative integer-valued random variable, then*

$$\mathbb{P}(X = 0) \leq \frac{\mathbb{V}\mathrm{ar}[X]}{(\mathbb{E}[X])^2} = \frac{\mathbb{E}[X^2]}{(\mathbb{E}[X])^2} - 1\,.$$

The second moment method can easily be strengthened using the Cauchy-Schwarz inequality.

Theorem 1.5.7 (The Cauchy-Schwarz inequality)**.** *For all vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, we have that*

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right).$$

Theorem 1.5.8 (The strong second moment method)**.** *If X is a nonnegative integer-valued random variable, then*

$$\mathbb{P}(X = 0) \leq \frac{\mathbb{V}\mathrm{ar}[X]}{\mathbb{E}[X^2]} = 1 - \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}\,.$$

Proof. Note that $X = X \cdot I_{X>0}$. But

$$\begin{aligned}(\mathbb{E}[X])^2 &= \big(\mathbb{E}[X \cdot I_{X>0}]\big)^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[I_{X>0}^2] \\ &= \mathbb{E}[X^2] \cdot \mathbb{E}[I_{X>0}] = \mathbb{E}[X^2] \cdot \mathbb{P}(X > 0)\end{aligned}$$

yields

$$\mathbb{P}(X = 0) = 1 - \mathbb{P}(X > 0) \leq 1 - \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]} = \frac{\mathbb{V}\mathrm{ar}[X]}{\mathbb{E}[X^2]}\,.$$

The last equality follows immediately from the definition of the variance. □

The bound in Theorem 1.5.8 is better than the bound in Theorem 1.5.6, since $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$. For many applications, however, these bounds are equally powerful.

Examples of Both Methods

We will be using both the first and second moment methods many times in this book, especially the former. For illustrative purposes, we conclude this section with simple examples of both methods. In particular, we will prove that $p = \log n/n$ is the threshold for the disappearance of isolated vertices in $\mathbb{G}(n, p)$.

First, we will show that a.a.s. there is no isolated vertex in a random graph $\mathbb{G}(n, p)$ for

$$p = p(n) = \frac{\ln n + \omega(n)}{n},$$

where $\omega(n)$ is any function tending to infinity. (The case when $\omega(n)$ tends to a constant $c \in \mathbb{R}$ will be discussed later.) Let I_i denote the indicator random variable for the event when the vertex i is isolated, where $i \in [n]$. The number of isolated vertices in $\mathbb{G}(n, p)$ is $X = \sum_{i=1}^{n} I_i$. Since for every $i \in [n]$

$$\mathbb{P}(I_i = 1) = (1-p)^{n-1} \sim \exp\big(-\ln n - \omega(n)\big) = \frac{e^{-\omega(n)}}{n},$$

$\mathbb{E}[X] \sim e^{-\omega(n)} \to 0$ as $n \to \infty$, and the claim holds by the first moment method.

Now, we will show that a.a.s. there is at least one isolated vertex in a random graph $\mathbb{G}(n, p)$ for

$$p = p(n) = \frac{\ln n - \omega(n)}{n}.$$

In this case, $\mathbb{E}[X] \sim e^{\omega(n)} \to \infty$ as $n \to \infty$, and

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X] &= \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{n} I_i\right] = \sum_{1\le i,j\le n} \mathbb{C}\mathrm{ov}(I_i, I_j) \\ &= \sum_{1\le i,j\le n} (\mathbb{E}[I_i I_j] - (\mathbb{E}[I_i])(\mathbb{E}[I_j])).\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X] &= \sum_{1\le i,j\le n, i\neq j} \Big(\mathbb{P}(I_i = 1, I_j = 1) - (\mathbb{P}(I_i = 1))^2\Big) \\ &\quad + \sum_{i=1}^{n} \Big(\mathbb{P}(I_i = 1) - (\mathbb{P}(I_i = 1))^2\Big).\end{aligned}$$

The second term in the last sum can be dropped to derive the following bound:

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X] &\le \sum_{1\le i,j\le n, i\neq j} \Big((1-p)^{2n-3} - (1-p)^{2n-2}\Big) + \mathbb{E}[X] \\ &= \sum_{1\le i,j\le n, i\neq j} (1-p)^{2n-3}(1-(1-p)) + \mathbb{E}[X] \\ &\sim \sum_{1\le i,j\le n, i\neq j} \frac{e^{2\omega(n)}}{n^2} p + \mathbb{E}[X] \sim e^{2\omega(n)} p + \mathbb{E}[X].\end{aligned}$$

Hence, from the second moment method we derive that

$$\mathbb{P}(X = 0) \le \frac{\mathbb{V}\mathrm{ar}[X]}{(\mathbb{E}[X])^2} \le (1 + o(1))p + \frac{1}{\mathbb{E}[X]} = o(1).$$

1.6 Tools: Chernoff Bounds

Suppose that S is a random variable and $t > 0$. We would like to find the upper and lower tails of the distribution; that is, bounds for $\mathbb{P}(S \ge \mathbb{E}[S]+t)$ and $\mathbb{P}(S \le \mathbb{E}[S]-t)$. Let $u \ge 0$. Then

$$\mathbb{P}(S \ge \mathbb{E}[S] + t) = \mathbb{P}(e^{uS} \ge e^{u(\mathbb{E}[S]+t)}) \le e^{-u(\mathbb{E}[S]+t)}\mathbb{E}[e^{uS}],$$

by Markov's inequality. Similarly, for $u \le 0$,

$$\mathbb{P}(S \le \mathbb{E}[S] - t) \le e^{-u(\mathbb{E}[S]-t)}\mathbb{E}[e^{uS}].$$

Combining these inequalities, we obtain a bound for $\mathbb{P}(|S - \mathbb{E}[S]| \geq t)$.

Now, let $S_n = \sum_{i=1}^{n} X_i$, where $X_i, i \in [n]$ are independent random variables. Then for $u \geq 0$, we have that

$$\mathbb{P}(S_n \geq \mathbb{E}[S_n] + t) \leq e^{-u(\mathbb{E}[S_n]+t)} \prod_{i=1}^{n} \mathbb{E}[e^{uX_i}],$$

whereas for $u \leq 0$, we have that

$$\mathbb{P}(S_n \leq \mathbb{E}[S_n] - t) \leq e^{-u(\mathbb{E}[S_n]-t)} \prod_{i=1}^{n} \mathbb{E}[e^{uX_i}].$$

After calculating $\mathbb{E}[e^{uX_i}]$ and finding the value of u that minimizes the right side, we derive the desired bound.

To illustrate this general approach, we focus on Bernoulli(p) random variables. Then S_n has a Bin(n, p) distribution with expectation $\mu = \mathbb{E}[S_n] = np$. For $u \geq 0$, we have that

$$\mathbb{P}(S_n \geq \mu + t) \leq e^{-u(\mu+t)} \big(pe^u + (1-p)\big)^n.$$

To minimize the right side, we take

$$e^u = \frac{(\mu + t)(1-p)}{(n-\mu-t)p}.$$

Hence, assuming that $\mu + t < n$,

$$\mathbb{P}(S_n \geq \mu + t) \leq \left(\frac{\mu}{\mu+t}\right)^{\mu+t} \left(\frac{n-\mu}{n-\mu-t}\right)^{n-\mu-t},$$

whereas for $\mu + t > n$ this probability is zero.

Now, let

$$\varphi(x) = \begin{cases} (1+x)\log(1+x) - x, & x > -1, \\ \infty, & \text{otherwise.} \end{cases}$$

For $0 \leq t < n - \mu$, we have that

$$\begin{aligned} \mathbb{P}(S_n \geq \mu + t) &\leq \exp\left(-\mu\varphi\left(\frac{t}{\mu}\right) - (n-\mu)\varphi\left(\frac{-t}{n-\mu}\right)\right) \\ &\leq e^{-\mu\varphi(t/\mu)}, \end{aligned}$$

since $\varphi(x) \geq 0$ for every x. By a similar argument, for $0 \leq t < \mu$ we obtain that

$$\begin{aligned} \mathbb{P}(S_n \leq \mu - t) &\leq \exp\left(-\mu\varphi\left(\frac{-t}{\mu}\right) - (n-\mu)\varphi\left(\frac{t}{n-\mu}\right)\right) \\ &\leq e^{-\mu\varphi(-t/\mu)}. \end{aligned}$$

Now, observe that $\varphi(0) = 0$ and

$$\varphi'(x) = \log(1+x) \leq x = \left(x^2/2\right)'.$$

Thus, $\varphi(x) \geq x^2/2$ for $-1 \leq x \leq 0$. Further, $\varphi'(0) = 0$ and

$$\varphi''(x) = \frac{1}{1+x} \geq \frac{1}{(1+x/3)^3} = \left(\frac{x^2}{2(1+x/3)}\right)'',$$

so for $x \geq 0$

$$\varphi(x) \geq \frac{x^2}{2(1+x/3)}.$$

The functions $\varphi(x)$ (brown), $\frac{x^2}{2}$ (in green), and $\frac{x^2}{2(1+x/3)}$ (in red) are presented in Figure 1.7.

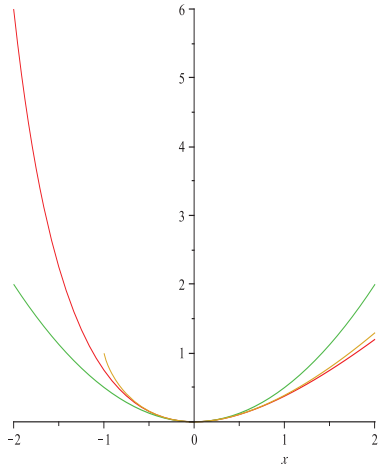

FIGURE 1.7: The functions: $\varphi(x)$, $\frac{x^2}{2}$, and $\frac{x^2}{2(1+x/3)}$.

Therefore, we arrive at the following result.

Theorem 1.6.1 (Chernoff bounds; see, for example, [115])**.** *If S_n is a random variable with the binomial distribution* $\mathrm{Bin}(n,p)$ *and $\mu = \mathbb{E}[S_n] = np$, then for $t \geq 0$ we have that*

$$\begin{aligned} \mathbb{P}(S_n \geq \mu + t) &\leq \exp\left(-\frac{t^2}{2(\mu + t/3)}\right) \text{ and} \\ \mathbb{P}(S_n \leq \mu - t) &\leq \exp\left(-\frac{t^2}{2\mu}\right). \end{aligned}$$

The following corollary is sometimes more convenient.

Corollary 1.6.2 ([115]). *If S_n is a random variable with the binomial distribution* $\mathrm{Bin}(n,p)$ *and $\mu = \mathbb{E}[S_n] = np$, then for $\varepsilon \leq 3/2$ we have that*

$$\mathbb{P}(|S_n - \mu| \geq \varepsilon\mu) \leq 2\exp\left(-\frac{\varepsilon^2}{3}\mu\right).$$

We mention another, sometimes useful version of the Chernoff bounds.

Theorem 1.6.3 ([8]). *If S_n is a random variable with the binomial distribution* $\mathrm{Bin}(n,p)$ *and $\mu = \mathbb{E}[S_n] = np$, then for $a > 0$ we have that*

$$\mathbb{P}(S_n > \mu + a) < e^{-2a^2/n} \quad \text{and} \quad \mathbb{P}(S_n < \mu - a) < e^{-2a^2/n}.$$

In addition, all of the above bounds hold for the general case in which $X_i \in \mathrm{Bernoulli}(p_i)$ with (possibly) different p_i. Indeed, we can repeat all calculations with the only difference being that now

$$\prod_{i=1}^{n} \mathbb{E}[e^{uX_i}] = \prod_{i=1}^{n} \big(p_i e^u + (1-p_i)\big).$$

We need the following inequality known as the *arithmetic-geometric mean inequality*.

Lemma 1.6.4. *For all sequences of nonnegative numbers $(a_1, a_2, \ldots, a_n)$ we have that*

$$\frac{1}{n}\sum_{i=1}^{n} a_i \geq \left(\prod_{i=1}^{n} a_i\right)^{1/n}.$$

Using the arithmetic-geometric mean inequality we derive that

$$\prod_{i=1}^{n} \mathbb{E}[e^{uX_i}] \leq \left(\frac{1}{n}\sum_{i=1}^{n} \big(p_i e^u + (1-p_i)\big)\right)^n = \big(pe^u + (1-p)\big)^n,$$

where $p = \sum_{i=1}^{n} p_i/n$. This is exactly the same expression as we had before with p taken as the arithmetic mean of the p_i's. The rest of the proof is not affected.

Finally, recall that the hypergeometric distribution describes the probability of k successes in n draws, without replacement, from a population of size N that contains exactly K successes, wherein each draw is either a success or a failure. Note that drawing with replacement would yield a binomial random variable. It seems reasonable that drawing without replacement tends to produce smaller random fluctuations, and indeed the bounds obtained above (Theorem 1.6.1 and Corollary 1.6.2) still hold for $Hypergeometric(N, K, n)$ with $\mu = nK/N$. For more details, see, for example [115].

Lower Bound

Until now, we have focused on bounding from above the probability that a random variable is far away from the expectation. However, sometimes we are more interested in bounding the probability of this rare event from below. The well-known Central Limit Theorem suggests that the distribution of the sum of many independent random variables is approximately normal, and so the bounds we obtained earlier should not be far from the truth. This is actually the case under general circumstances.

Theorem 1.6.5. *If S_n is a random variable with the binomial distribution* $\mathrm{Bin}(n, 1/2)$ *and $\mathbb{E}[S_n] = n/2 = \mu$, then for any integer $t \in [0, n/8]$ we have that*

$$\mathbb{P}(S_n \geq \mu + t) \geq \frac{1}{15} e^{-16t^2/n}.$$

Such general and precise bounds can be found in [85]. We use some elementary calculations to prove Theorem 1.6.5.

Proof of Theorem 1.6.5. We note that

$$\begin{aligned}
\mathbb{P}(S_n \geq n/2 + t) &= \sum_{j=t}^{n/2} \mathbb{P}(S_n = n/2 + j) \\
&= 2^{-n} \sum_{j=t}^{n/2} \binom{n}{n/2 + j} \geq 2^{-n} \sum_{j=t}^{2t-1} \binom{n}{n/2 + j} \\
&= 2^{-n} \sum_{j=t}^{2t-1} \binom{n}{n/2} \frac{n/2}{n/2 + j} \cdot \frac{n/2 - 1}{n/2 + j - 1} \\
&\quad \cdots \frac{n/2 - j + 1}{n/2 + 1}.
\end{aligned}$$

As $\binom{n}{n/2} \geq 2^n/2\sqrt{n/2} = 2^n/\sqrt{2n}$ (see Lemma 1.3.2), we derive that

$$\begin{aligned}
\mathbb{P}(S_n \geq n/2 + t) &\geq \frac{1}{\sqrt{2n}} \sum_{j=t}^{2t-1} \prod_{i=1}^{j} \left(1 - \frac{j}{n/2+i}\right) \\
&\geq \frac{t}{\sqrt{2n}} \left(1 - \frac{2t}{n/2}\right)^{2t} \\
&\geq \frac{t}{\sqrt{2n}} \exp\left(-\frac{16t^2}{n}\right),
\end{aligned}$$

since $1 - x \geq e^{-2x}$ for $0 \leq x \leq 1/2$. For $t \geq \frac{1}{4\sqrt{2}}\sqrt{n}$, the probability is at least $\frac{1}{8}e^{-16t^2/n}$. For $0 \leq t < \frac{1}{4\sqrt{2}}\sqrt{n}$, we have that

$$\begin{aligned}
\mathbb{P}(S_n \geq n/2 + t) &\geq \mathbb{P}\left(S_n \geq n/2 + \frac{1}{4\sqrt{2}}\sqrt{n}\right) \\
&\geq \frac{1}{8}e^{-1/2} \geq \frac{1}{15} \geq \frac{1}{15}e^{-16t^2/n},
\end{aligned}$$

and so the claimed bound also holds for this range of t. □