Finite Probability Spaces
Lecture Notes

László Babai

April 18, 1999

1 Finite Probability Spaces and Events

Definition 1.1 *A* **finite probability space** *is a finite set* $\Omega \neq \emptyset$ *together with a function* $\Pr : \Omega \to \mathbf{R}^+$ *such that*

1. $\forall \omega \in \Omega,\ \Pr(\omega) > 0$
2. $\sum_{\omega \in \Omega} \Pr(\omega) = 1.$

The set Ω is the **sample space** and the function Pr is the **probability distribution**. The elements $\omega \in \Omega$ are called **atomic events** or **elementary events**. An **event** is a subset of Ω. For $A \subseteq \Omega$, we define the **probability** of A to be $\Pr(A) := \sum_{\omega \in A} \Pr(\omega)$. In particular, for atomic events we have $\Pr(\{\omega\}) = \Pr(\omega)$; and $\Pr(\emptyset) = 0$, $\Pr(\Omega) = 1$. The **trivial events** are those with probability 0 or 1, i. e. $\emptyset$ and Ω.

The **uniform distribution** over the sample space Ω is defined by setting $\Pr(\omega) = 1/|\Omega|$ for every $\omega \in \Omega$. With this distribution, we shall speak of the **uniform probability space** over Ω. In a uniform space, calculation of probabilities amounts to counting: $\Pr(A) = |A|/|\Omega|$.

Exercise 1 *In the card game of bridge, a deck of 52 cards are evenly distributed among four players called North, East, South, and West. What sample space does each of the following questions refer to: (a) What is the probability that North holds all the aces? (b) What is the probability that each player holds one of the aces? – These questions refer to uniform probability spaces. Calculate the probabilities.*

Observation 1.2 $\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B)$.

Definition 1.3 *Events A and B are* **disjoint** *if* $A \cap B = \emptyset$.

Consequence 1.4 $\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^{k} \Pr(A_i)$, *and equality holds if and only if the* A_i *are pairwise disjoint.*

Definition 1.5 *If* A *and* B *are events and* $\Pr(B) > 0$ *then the* **conditional probability of** A **relative to** B, *written* $\Pr(A|B)$, *is given by* $\Pr(A|B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$.

We note that B can be viewed as a sample space with the probabilities being the conditional probabilities under condition B.

Note that $\Pr(A \cap B) = \Pr(A|B)\Pr(B)$.

Exercise 2 *Prove:* $\Pr(A \cap B \cap C) = \Pr(A|B \cap C)\Pr(B|C)\Pr(C)$.

Exercise 3 *We roll three dice. What is the probability that the sum of the three numbers we obtain is 9? What is the probability that the first die shows 5? What is the conditional probability of this event assuming the sum of the numbers shown is 9? – What is the probability space in this problem? How large is the sample space?*

A **partition** of Ω is a family of pairwise disjoint events $H_1, \ldots, H_m$ covering Ω:

$$\Omega = H_1 \cup \ldots \cup H_k, \qquad H_i \cap H_j = \emptyset. \tag{1}$$

We usually assume that each event in the partition is nontrivial.

Exercise 4 *Prove: given a partition* $(H_1, \ldots, H_k)$ *of* Ω, *we have*

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A|H_i)\Pr(H_i). \tag{2}$$

for any event A.

The significance of this formula is that the conditional probabilities are sometimes easier to calculate than the left hand side.

Definition 1.6 *Events* A *and* B *are* **independent** *if* $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Exercise 5 *If* $\Pr(B) > 0$ *then:* A *and* B *are independent* $\Longleftrightarrow \Pr(A|B) = \Pr(A)$.

Exercise 6 *Prove: if* A *and* B *are independent events then* $\bar{A}$ *and* B *are also independent, where* $\bar{A} = \Omega \setminus A$.

Exercise 7 *(a) If we roll a die, are the following events independent: "the number shown is odd"; "the number shown is prime"? (b) Let us consider a uniform probability space over a sample space whose cardinality is a prime number. Prove that no two non-trivial events can be independent.*

Note that the trivial events are independent of any other events, i. e. if a trivial event is added to a collection of independent events, they remain independent.

The events A and B are said to be **positively correlated** if $\Pr(A \cap B) > \Pr(A)\Pr(B)$. They are **negatively correlated** if $\Pr(A \cap B) < \Pr(A)\Pr(B)$.

Exercise 8 *Are the two events described in Exercise 3 positively, or negatively correlated, or independent?*

Exercise 9 *Prove: two events A, B are positively (negatively) correlated if and only if* $\Pr(B|A) > \Pr(B)$ *(*$\Pr(B|A) < \Pr(B)$*, resp.).*

Definition 1.7 *Events $A_1, \ldots, A_k$ are* **independent** *if for all subsets $S \subseteq \{1, \ldots, k\}$, we have*

$$\Pr(\cap_{i \in S} A_i) = \prod_{i \in S} \Pr(A_i). \tag{3}$$

Note that if $k \geq 3$, then the statement that events $A_1, \ldots, A_k$ are independent is stronger than pairwise independence. For example, pairwise independence does not imply triplewise independence. For added emphasis, independent events are sometimes called *fully* independent, or *mutually* independent, or *collectionwise* independent.

Exercise 10 *Construct 3 events which are pairwise independent but not collectionwise independent. What is the smallest sample space for this problem?*

(See the end of this section for more general problems of this type.)

Exercise 11 *Prove: if the events A, B, C, D, E are independent then the events $A \setminus B$, $C \cup D$, and E are independent as well. Formulate a general statement, for n events grouped into blocks.*

Exercise 12 *We have n balls colored red, blue, and green (each ball has exactly one color and each color occurs at least once). We select k of the balls with replacement (independently, with uniform distribution). Let A denote the event that the k balls selected have the same color. Let p_r denote the conditional probaility that the first ball selected is red, assuming condition A. Define p_b and p_g analogously for blue and green outcomes. Assume $p_1 + p_2 = p_3$. Prove: $k \leq 2$. Show that $k = 2$ is actually possible.*

Exercise 13 *(Random graphs) Consider the uniform probability space over the set of all the* $2^{\binom{n}{2}}$ *graphs with a given set* V *of* n *vertices. (a) What is the probability that a particular pair of vertices is adjacent? Prove that these* $\binom{n}{2}$ *events are independent. (b) What is the probability that the degrees of vertex 1 and vertex 2 are equal? Give a closed-form expression. (c) If* p_n *denotes the probability calculated in part (b), prove that* $p_n\sqrt{n}$ *tends to a finite positive limit and determine its value. (c) How are the following two events correlated: A:"vertex 1 has degree 3"; B:"vertex 2 has degree 3"? Asymptotically evaluate the ratio* $\Pr(A|B)/\Pr(A)$.

In exercises like the last one, one often has to estimate binomial coefficients. The following result comes handy:

Stirling's formula.

$$n! \sim (n/e)^n\sqrt{2\pi n}. \tag{4}$$

Here the $\sim$ notation refers to asymptotic equality: for two sequences of numbers a_n and b_n we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. While such formulae are excellent in predicting what happens for "large" n, they do not tell how large is large enough.

A stronger, non-asymptotic variant, giving useful results for specific values of n, is the following:

$$n! = (n/e)^n\sqrt{2\pi n}(1 + \theta_n/(12n)), \tag{5}$$

where $|\theta_n| \leq 1$.

Exercise 14 *Evaluate asymptotically the binomial coefficient* $\binom{2n}{n}$. *Show that* $\binom{2n}{n} \sim c\cdot 4^n/\sqrt{n}$ *where* c *is a constant. Determine the value of* c.

We mention some important asymptotic relations from number theory. Let $\pi(x)$ denote the number of all prime numbers $\leq x$, so $\pi(2) = 1$, $\pi(10) = 4$, etc. The **Prime Number Theorem** of Hadamard and de la Vallée-Poussin asserts that

$$\pi(x) \sim x/\ln x. \tag{6}$$

Another important relation estimates the sum of reciprocals of prime numbers. The summation below extends over all primes $p \leq x$.

$$\sum_{p\leq x} 1/p \sim \ln\ln x. \tag{7}$$

In fact a stronger result holds: there exists a number B such that

$$\lim_{x\to\infty}\left(\sum_{p\leq x} 1/p - \ln\ln x\right) = B. \tag{8}$$

(Deduce (7) from (8).)

Exercise 15 *Assuming 100-digit integers are "large enough" for the Prime Number Theorem to give a good approximation, estimate the probability that a random integer with at most 100 decimal digits is prime. (The integer is drawn with uniform probability from all positive integers in the given range.)*

Exercise 16 *Construct a sample space Ω and events $A_1, \ldots, A_n$ $(\forall n \geq 2)$ such that* $\Pr(A_i) = 1/2$*, every $n-1$ of the A_i are independent, but the n events are* not *independent.*

Exercise 17 *(*) Let $1 \leq k \leq n-1$. (a) Construct a sample space Ω and n events such that every k of these n events are independent; but no $k+1$ of these events are independent. (b) Solve part (a) under the additional constraint that each of the n events have probability 1/2.*

(Hint. Take a k-dimensional vector space W over a finite field of order $q \geq n$. Select n vectors from W so that any k are linearly independent. Let W be the sample space.)

Exercise 18 *Suppose we have n independent nontrivial events. Prove: $|\Omega| \geq 2^n$.*

Exercise 19 *(Small sample space for pairwise independent events.) (a) For $n = 2^k - 1$, construct a probability space of size $n+1$ with n pairwise independent events each of probability $1/2$. (b)* Same for n a prime number of the form $n = 4k - 1$.*

Exercise 20 *(*) Prove: if there exist n pairwise independent nontrivial events in a probability space then $|\Omega| \geq n+1$. (If this is too difficult, solve the special case when all events considered have probability 1/2 and the space is uniform.)*

2 Random Variables and Expected Value

Definition 2.1 *A* **random variable** *is a function $\xi : \Omega \to \mathbf{R}$.*

We say that ξ is **constant** if $\xi(\omega)$ takes the same value for all $\omega \in \Omega$.

Definition 2.2 *The* **expected value** *of a random variable ξ is* $\mathrm{E}(\xi) = \sum_{\omega \in \Omega} \xi(\omega) \Pr(\omega)$.

Proposition 2.3 *Let $\{u_1, \ldots, u_k\}$ be the set of (distinct) values taken by ξ. Let $p_i = \Pr(\xi = u_i)$, where the statement "$\xi = u_i$" refers to the event $\{\omega : \xi(\omega) = u_i\}$. Then* $\mathrm{E}(\xi) = \sum_{i=1}^k u_i p_i$.

Proof: Exercise.

Exercise 21

$$\min \xi \leq \mathrm{E}(\xi) \leq \max \xi. \tag{9}$$

Throughout these notes, $\xi, \eta, \zeta, \vartheta$, and their subscripted versions refer to random variables.

Proposition 2.4 (Additivity of the Expected Value) *Let $\xi_1, \ldots, \xi_k$ be arbitrary random variables. Then*

$$\mathrm{E}(\xi_1 + \cdots + \xi_k) = \sum_{i=1}^{k} \mathrm{E}(\xi_i) \tag{10}$$

Proof: $\mathrm{E}(\sum_{i=1}^{k} \xi_i) = \sum_{\omega \in \Omega} (\xi_1(\omega) + \cdots + \xi_k(\omega)) \Pr(\omega) = \sum_{i=1}^{k} \sum_{\omega \in \Omega} \xi_i \Pr(\omega) = \sum_{i=1}^{k} \mathrm{E}(\xi_i).$ ■

Exercise 22 *(Linearity of the expected value.) If $c_1, \ldots, c_k$ are constants then*

$$\mathrm{E}(\sum_{i=1}^{k} c_i \xi_i) = \sum_{i=1}^{k} c_i \mathrm{E}(\xi_i). \tag{11}$$

Definition 2.5 *The* **indicator variable** *of an event $A \subseteq \Omega$ is the function $\vartheta_A : \Omega \to \{0, 1\}$ given by*

$$\vartheta_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A \end{cases}$$

Exercise 23 *The expected value of an indicator variable is $\mathrm{E}(\vartheta_A) = \Pr(A)$.*

Indicator variables are particularly useful if we want to count events. Some of the exercises at the end of this section should serve as examples.

Exercise 24 *(a) Every random variable ξ is a linear combination of indicator variables. (b) Given a random variable ξ there exist functions $f_1, \ldots, f_k$ such that the random variables $\xi_i := f_i(\xi)$ are indicator variables and ξ is a linear combination of the ξ_i.*

We say that ξ is **nonnegative** if $\xi(\omega) \geq 0$ for all $\omega \in \Omega$.

Theorem 2.6 *(***Markov's Inequality***): If ξ is nonnegative then $\forall a > 0$,*

$$Pr(\xi \geq a) \leq \frac{\mathrm{E}(\xi)}{a}.$$

Proof: Let $m = \mathrm{E}(\xi) > 0$. Then $m = \sum_i \mu_i \Pr(\xi = \mu_i) \geq$
$\geq \sum_{\mu_i \geq a} \mu_i \Pr(\xi = \mu_i)$ (we just omitted some terms; all terms are nonnegative)
$\geq a \sum_{\mu_i \geq a} \Pr(\xi = \mu_i) = a \Pr(\xi \geq a)$ (sum of disjoint events).
So we have $m \geq a \Pr(\xi \geq a)$. ∎

Exercise 25 *Suppose in a lottery you have to pick five different numbers from 1 to 90. Then five winning numbers are drawn. If you picked two of them, you win 20 dollars. For three, you win 150 dollars. For four, you win 5,000 dollars, and if all the five match, you win a million. (a) What is the probability that you picked exactly three of the winning numbers? (b) What is your expected win? (c) What does Markov's inequality predict about the probability that you'll win at least 20 dollars? (d) What is the actual probability that this happens?*

Exercise 26 *A club with 2000 members distributes membership cards numbered 1 through 2000 to its members at random; each of the 2000! permutations of the cards is equally likely. Members whose card number happens to coincide with their year of birth receive a prize. Determine the expected number of lucky members.*

Exercise 27 *What is the expected number of edges in a random graph? What is the expected number of triangles? (There are n vertices; each pair is adjacent with probability 1/2 independently.)*

Exercise 28 *Let n be a random integer, chosen uniformly between 1 and N. What is the expected number of distinct prime divisors of n? Show that the result is asymptotically equal to* $\ln \ln N$ *(as $N \to \infty$).*

3 Standard deviation and Chebyshev's Inequality

Definition 3.1 *The k^{th}* **moment** *of ξ is* $\mathrm{E}(\xi^k)$. *The k^{th}* **central moment** *of ξ is the k^{th} moment of $\xi - \mathrm{E}(\xi)$, i.e.* $\mathrm{E}((\xi - \mathrm{E}(\xi))^k)$.

Definition 3.2 *The* **variance** *of ξ is its second central moment,* $\mathrm{Var}(\xi) := \mathrm{E}((\xi - \mathrm{E}(\xi))^2)$.

Note that the variance is always nonnegative. It is zero exactly if ξ is constant. (Why?)

Definition 3.3 *The* **standard deviation** *of ξ is* $\sigma(\xi) := \sqrt{\mathrm{Var}(\xi)}$.

Exercise 29 *(Invariance under shifts.) Prove that if c is a constant then* $\mathrm{Var}(\xi) = \mathrm{Var}(\xi + c)$*; and consequently, $\sigma(\xi) = \sigma(\xi + c)$.*

Exercise 30 *Prove: if c is a constant then* $\mathrm{Var}(\mathrm{c}\xi) = \mathrm{c}^2\mathrm{Var}(\xi)$*; and consequently,* $\sigma(c\xi) = |c|\sigma(\xi)$.

Observation 3.4 $\mathrm{Var}(\xi) = \mathrm{E}(\xi^2) - (\mathrm{E}(\xi))^2$.

Corollary 3.5 (Cauchy-Schwarz inequality) $(\mathrm{E}(\xi))^2 \le \mathrm{E}(\xi^2)$. ■

Proof of Observation: Let $m = \mathrm{E}(\xi)$. Then $\mathrm{Var}(\xi) = \mathrm{E}((\xi - m)^2) = \mathrm{E}(\xi^2 - 2\xi m + m^2) = \mathrm{E}(\xi^2) - 2m\mathrm{E}(\xi) + \mathrm{E}(m^2) = \mathrm{E}(\xi^2) - 2mm + m^2 = \mathrm{E}(\xi^2) - m^2$. ■

Chebyshev's inequality tells us that random variables don't like to stray away from their expected value by more than a small multiple of their standard deviation.

Theorem 3.6 *(***Chebyshev's Inequality***): Let $m = \mathrm{E}(\xi)$. Then for any number $a > 0$,*

$$\Pr(|\xi - m| \ge a) \le \frac{\mathrm{Var}(\xi)}{a^2}. \tag{12}$$

Proof: Let $\eta = (\xi - m)^2$. Then, by definition, $\mathrm{E}(\eta) = \mathrm{Var}(\xi)$. We apply Markov's Inequality to the nonnegative random variable η: $\Pr(|\xi - m| \ge a) = \Pr(\eta \ge a^2) \le \mathrm{E}(\eta)/a^2 = \mathrm{Var}(\xi)/a^2$. ■

Exercise 31 *In its more common form the Cauchy-Schwarz inequality asserts that for any real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$ we have*

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \ge \left(\sum_{i=1}^{n} x_i y_i\right)^2. \tag{13}$$

Deduce this inequality from Corollary 3.5.

Exercise 32 *(Limit on negatively correlated events.) Suppose the events $A_1, \ldots, A_m$ each have probability $1/2$ and for each i, j, $\Pr(|A_i \cap A_j| \le 1/5$. Prove: $m \le 6$. Generalize the statement to events of probability p, with $p^2 - \epsilon$ in the place of $1/5$.*

Exercise 33 *Prove: if the k^{th} moment of ξ is zero for all odd integers $k > 0$ then $\Pr(\xi = u) = \Pr(\xi = -u)$ for all $u \in \mathbf{R}$.*

4 Independence of random variables

Definition 4.1 *$\xi_1, \ldots, \xi_k$ are* **independent** *if $\forall u_1, \ldots, u_k$,*

$$\Pr(\xi_1 = u_1, \ldots, \xi_k = u_k) = \prod_{i=1}^{k} \Pr(\xi_i = u_i). \tag{14}$$

Exercise 34 *Prove that the events $A_1, \ldots, A_k$ are independent if and only if their indicator variables are independent.*

Exercise 35 *Prove that the random variables $\xi_1, \ldots, \xi_k$ are independent if and only if for all choices of the numbers $u_1, \ldots, u_k$, the k events $\xi_1 = u_1, \ldots, \xi_k = u_k$ are independent. Show that this is also equivalent to the independence of all k-tuples of events of the form $\xi_1 < u_1, \ldots, \xi_k < u_k$.*

Exercise 36 *Prove: if $\xi_1, \ldots, \xi_k$ are independent then $f_1(\xi_1), \ldots, f_k(\xi_k)$ are also independent, where the f_i are arbitrary functions. For example, ξ_1^2, e^{ξ_2}, and $\cos(\xi_3)$ are independent.*

Exercise 37 *Prove: if ξ, η, ζ are independent random variables then $f(\xi, \eta)$ and ζ are also independent, where f is an arbitrary function. (For instance, $\xi + \eta$ and ζ, or $\xi\eta$ and ζ are independent.) Generalize this statement to several variables, grouped into blocks, and a function applied to each block.*

Theorem 4.2 (Multiplicativity of the expected value) *If $\xi_1, \ldots, \xi_m$ are independent, then*

$$\mathrm{E}(\prod_{i=1}^{m} \xi_i) = \prod_{i=1}^{m} \mathrm{E}(\xi_i). \tag{15}$$

Exercise 38 *Prove this result for indicator variables.*

Exercise 39 *Prove: if ξ, η are independent, then one can write ξ as a sum $\xi = c_1\xi_1 + \ldots + c_k\xi_k$ and η as $\eta = d_1\eta_1 + \ldots + d_\ell\eta_\ell$ where the ξ_i and η_j are indicator variables and for every i, j, the variables ξ_i and η_j are independent.*

Exercise 40 *Combine the two preceding exercises to a proof of the Theorem for $m = 2$ variables.*

Exercise 41 *Deduce the general case from the preceding exercise by induction on m, using Exercise 37.*

This sequence completes the proof of Theorem 4.2. ∎

While this result required the full force of independence of our random variables, in the next result, only pairwise independence is required.

Theorem 4.3 (Additivity of the variance.) *Let* $\eta = \xi_1 + \xi_2 + \cdots + \xi_k$. *If* $\xi_1, \ldots, \xi_k$ *are* pairwise independent *then* $\mathrm{Var}(\eta) = \sum_{i=1}^k \mathrm{Var}(\xi_i)$.

Proof: By Exercise 29, we may assume that $\mathrm{E}(\xi_i) = 0$ (otherwise we replace each ξ_i by $\xi_i - \mathrm{E}(\xi_i)$; this will not change the variance, nor does it affect independence (why?)). Having made this assumption it follows that $\mathrm{E}(\eta) = 0$. Moreover, for $i \neq j$ we have $\mathrm{E}(\xi_i\xi_j) = \mathrm{E}(\xi_i)\mathrm{E}(\xi_j) = 0$ by pairwise independence.

It follows that $\mathrm{Var}(\xi_i) = \mathrm{E}(\xi_i^2)$ and $\mathrm{Var}(\eta) = \mathrm{E}(\eta^2) = \mathrm{E}((\sum \xi_i)^2) = \mathrm{E}(\sum_i \xi_i^2 + 2\sum_{i<j} \xi_i\xi_j) = \sum_i \mathrm{E}(\xi_i^2) + 2\sum_{i<j} \mathrm{E}(\xi_i\xi_j) = \sum_i \mathrm{Var}(\xi_i)$. ∎

Corollary 4.4 *Let* $\xi_1, \ldots, \xi_n$ *be random variables with the same standard deviation* σ. *Let us consider their average,* $\eta := (1/n)\sum_{i=1}^n \xi_i$. *If the* ξ_i *are pairwise independent then* $\sigma(\eta) = \sigma/\sqrt{n}$. ∎

Corollary 4.5 (Weak law of large numbers.) *Let* $\xi_1, \xi_2, \ldots$ *be an infinite sequence of pairwise independent random variables each with expected value* m *and standard deviation* σ. *Let* $\eta_n = (1/n)\sum_{i=1}^n \xi_i$. *Then for any* $\delta > 0$,

$$\lim_{n\to\infty} \Pr(|\eta_n - m| > \delta) = 0. \tag{16}$$

Proof: Use Chebyshev's inequality and the preceding corollary. We obtain that the probability in question is $\leq \sigma^2/(\delta n) \to 0$ (as $n \to \infty$). ∎

Remark 4.6 Strictly speaking, we bent our rules here. An infinite sequence of non-constant, pairwise independent variables requires an infinite sample space. What we actually proved, then, is the following. Let us fix the values m and $\sigma \geq 0$. Assume that we are given an infinite sequence of finite probability spaces, and over the n^{th} space, we are given n independent random variables $\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,n}$. Let $\eta_n = (1/n)\sum_{i=1}^n \xi_{n,i}$. Then for any $\delta > 0$, the limit relation (16) holds.

5 Chernoff's Bound

Although the bound in the proof of the Weak Law of Large Numbers tends to zero, it does so rather slowly. If our variables are fully independent and bounded, much stronger estimates can be obtained by a method due to Chernoff. The bounds will go to zero exponentially as a function of n, and this is what most combinatorial applications require.

For example, let us consider a sequence of n independent coin flips; let ψ denote the number of heads in this sequence. Then $\mathrm{E}(\psi) = n/2$ and $\mathrm{Var}(\xi) = n/4$ (by the additivity of the variance). Therefore Chebyshev's inequality tells us that

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < \frac{1}{4r^2}. \tag{17}$$

Below we shall prove the much stronger inequality

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < 2e^{-2r^2}. \tag{18}$$

under the same conditions.

The following corollary illustrates the power of inequality (18).

Corollary 5.1 *For any $\varepsilon > 0$, almost all graphs have no vertices of degree $< (1-\varepsilon)n/2$ or $> (1+\varepsilon)n/2$ where n is the number of vertices.*

Proof of the Corollary. Let $V = \{1, \ldots, n\}$ be the vertex set of our random graph. Let δ_i denote the degree of vertex i; so δ_i is the number of heads in a sequence of $(n-1)$ independent coin flips. Therefore, by inequality (18), we have that

$$\Pr(|\delta_i - (n-1)/2| \geq r\sqrt{n-1}) < 2e^{-2r^2}. \tag{19}$$

Let us now set $r = \varepsilon\sqrt{n-1}$. Then we obtain

$$\Pr(|\delta_i - (n-1)/2| \geq \varepsilon(n-1)) < 2e^{-2\varepsilon^2(n-1)}. \tag{20}$$

Therefore the probability that there exists an i such that $|\delta_i - (n-1)/2| \geq \varepsilon(n-1)$ is less than n times the right hand side, i. e., less than $2ne^{-2\varepsilon^2(n-1)}$. This quantity approaches zero at an exponential rate as $n \to \infty$.

The slight change in the statement (having changed n to $n-1$) can be compensated for by slightly reducing ε. ∎

Note that the same procedure using inequality (17) will fail. Indeed, setting $r = \varepsilon\sqrt{n-1}$ in inequality (17), the right hand side will be $1/(4\varepsilon^2(n-1))$, and if we multiply this quantity by n, the result will be greater than 1 (if $\varepsilon < 1/2$, a meaningless upper bound for a probability.

Now we turn to the proof of inequality (18). It will be convenient to state the main result in terms of random variables with zero expected value.

Theorem 5.2 (Chernoff) *Let ξ_i be independent random variables satisfying* $\Pr(\xi_i = 1) = \Pr(\xi_i = -1) = 1/2$. *Let $\eta = \sum_{i=1}^n \xi_i$. Then for any $a > 0$,*

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{21}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{22}$$

Exercise 42 *Deduce inequality (18) from this theorem.*

Hint. Represent ψ as $\sum_{i=1}^n \theta_i$ where θ_i is the indicator variable of the i-th coin flip. Set $\xi_i = 2\theta_i - 1$ and $\eta = \sum_{i=1}^n \xi_i$. Note that $\psi - n/2 = \eta/2$. Apply Theorem 5.2 to the ξ_i and translate the result back to ψ.

Now we turn to the proof of Theorem 5.2.

Let t be a positive real number. We shall later suitably choose the value of t. Let us consider the random variables $\zeta_i := \exp(t\xi_i)$. (Notation: $\exp(x) = e^x$.) The ζ_i are again independent (for any fixed t) by Exercise 36. Therefore we can apply the multiplicativity of the expected value to them:

$$\mathrm{E}(e^{t\eta}) = \mathrm{E}(\exp(\sum_{i=1}^n t\xi_i)) = \mathrm{E}(\prod_{i=1}^n \zeta_i) == \prod_{i=1}^n \mathrm{E}(\zeta_i) = \prod_{i=1}^n \mathrm{E}(\exp(t\xi_i)). \tag{23}$$

Applying Markov's inequality to the variable $e^{t\eta}$, we conclude that

$$\Pr(\eta \geq a) = \Pr(e^{t\eta} \geq e^{ta}) \leq \prod_{i=1}^n \mathrm{E}(\exp(t\xi_i))e^{-ta}. \tag{24}$$

Recall that $\cosh(x) = (e^x + e^{-x})/2$ and observe that

$$\mathrm{E}(\exp(t\xi_i)) = \cosh(t). \tag{25}$$

Therefore the preceding inequality implies that

$$\Pr(\eta \geq a) < \frac{\cosh(t)^n}{e^{ta}}. \tag{26}$$

This is true for every $t > 0$. All we need to do is choose t appropriately to obtain the strongest possible result. To this end we need the following simple observation.

Lemma 5.3

$$\cosh(x) \leq e^{x^2/2}.$$

Proof: Compare the Taylor series of the two sides. On the left hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \dots \tag{27}$$

On the right hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \dots . \tag{28}$$

■

Consequently, from inequality (26) we infer that

$$\Pr(\eta \geq a) < \exp(t^2 n/2 - ta). \tag{29}$$

The expression $t^2 n/2 - ta$ is minimized when $t = a/n$; setting $t := a/n$ we conclude that $\Pr(\eta \geq a) < \exp(-a^2/2n)$, as required.

Replacing each ξ_i by $-\xi_i$ we obtain the inequality $\Pr(\eta \leq -a) < \exp(-a^2/2n)$; adding this to the preceding inequality we obtain $\Pr(|\eta| \leq a) < 2\exp(-a^2/2n)$. ■

We note that Chernoff's technique works under much more general circumstances. We state a useful and rather general case, noting that even this result does not exploit the full power of the method.

Theorem 5.4 (Chernoff) *Let ξ_i be independent random variables satisfying $|\xi_i| \leq 1$ and $\mathrm{E}(\xi_i) = 0$. Let $\eta = \sum_{i=1}^{n} \xi_i$. Then for any $a > 0$,*

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{30}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{31}$$

Proof: As before, we set $t = a/n$. Let

$$h(x) = \cosh(t) + x \cdot \sinh(t). \tag{32}$$

(Recall that $\sinh(t) = (e^t - e^{-t})/2$.) Observe that $h(x) \geq e^{tx}$ for all x in the interval $-1 \leq x \leq 1$. (The graph of $h(x)$ over the interval $[-1, 1]$ is the segment connecting the corresponding two points of the graph of the function e^{tx}, and e^{tx} is a convex function.)

Moreover, because of the linearity of the $h(x)$ function, we have $\mathrm{E}(h(\xi_i)) = h(E(\xi_i)) = h(0) = \cosh(t)$. Therefore

$$\mathrm{E}(e^{t\xi_i}) \leq \mathrm{E}(h(\xi_i)) = \cosh(t). \tag{33}$$

From here on the proof is identical with the proof of Theorem chernoff. ■