Day 1: The Probabilistic Method

Exercise 1. (Russia 1999)

In a certain school, every boy likes at least one girl. Prove that we can find a set S of at least half the students in the school such that each boy in S likes an odd number of girls in S.

Walkthrough:

(a) Flip a coin for every girl to determine whether she goes in S or not. What is the expected number of girls in S?
(b) Put every boy who likes an odd number of girls in S into S. What is the expected number of boys in S?
(c) What is the expected size of S? Conclude.

Exercise 2. Show that any graph G with m edges has a bipartite subgraph with $\geq \frac{m}{2}$ edges.

Walkthrough:

(a) Flip a coin on every vertex and define a corresponding bipartite subgraph.
(b) Show that the expected number of edges in the subgraph is $\frac{m}{2}$, and conclude.

Exercise 3. (USAMO 2012, problem 2)

A circle is divided into 432 congruent arcs by 432 points. The points are colored in four colors such that some 108 points are colored Red, some 108 points are colored Green, some 108 points are colored Blue, and the remaining 108 points are colored Yellow. Prove that one can choose three points of each color in such a way that the four triangles formed by the chosen points of the same color are congruent.

Walkthrough:

(a) Consider a random symmetry of the 432-gon formed by the points. How many are there? (Don't include the identity.)
(b) Color the red points that land on green points orange. What's the expected number of orange points? How many orange points are guaranteed to achievable?
(c) Modify and repeat step (b).
(d) Conclude.

Exercise 4. (Erdős)

Let $R(s)$ denote the Ramsey number of s, i.e., the smallest integer n for which, when one colors the edges of K_n either red or blue, there must be a monochromatic K_s.

Show that $R(s) > 2^{s/2}$ for $s \geq 3$.

Walkthrough:

(a) Let $n = \lfloor 2^{s/2} \rfloor$. Showing that $R(s) > n$ is showing that there existing a coloring of K_n with no monochromatic K_s. Randomly color each edge of K_n red or blue. What is the probability that a given set of s vertices forms a monochromatic K_s?
(b) Show that it suffices to show $\binom{n}{s} < 2^{\binom{s}{2}-1}$, and verify this is true by showing that $\binom{n}{s} < \frac{n^s}{2^s} < 2^{\binom{s}{2}-1}$.

Bonus Homework 5. Alice marks ten points in the plane. Is it always possible for Bob to place ten unit discs to cover all ten points, so that no two discs overlap?

Theorem 6 (Markov's Inequality)**.** *If $X \geq 0$ is a random variable and $a > 0$,*

$$P(X \geq a) \leq \frac{1}{a} \mathrm{E}[X],$$

with equality iff $X \in \{0, a\}$.

Homework 7. Prove this.

Definition 8. The covariance of two variables X and Y is

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

Homework 9. Show that $\mathrm{Cov}(X, Y)$ can also be written as $\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.

Definition 10. The variance of a random variable X is

$$\mathrm{Var}(X) := \mathrm{Cov}(X, X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2.$$

Theorem 11 (Chebyshev's Inequality)**.** *If X is a random variable and $a > 0$,*

$$P(|X - \mathrm{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2},$$

with equality iff $X - \mathrm{E}[X] \in \{0, a, -a\}$.

Proof. This is the direct result of plugging in $(X - \mathrm{E}[X])^2$ for X and a^2 for a into Markov's Inequality (**6**). □

Exercise 12. (USAMO 2012, problem 6)

For integer $n \geq 2$, let $x_1, x_2, \ldots, x_n$ be real numbers satisfying

$$x_1 + x_2 + \ldots + x_n = 0, \qquad \text{and} \qquad x_1^2 + x_2^2 + \ldots + x_n^2 = 1.$$

For each subset $A \subseteq \{1, 2, \ldots, n\}$, define

$$S_A = \sum_{i \in A} x_i.$$

(If A is the empty set, then $S_A = 0$.)

Prove that for any positive number λ, the number of sets A satisfying $S_A \geq \lambda$ is at most $2^{n-3}/\lambda^2$. For which choices of $x_1, x_2, \ldots, x_n, \lambda$ does equality hold?

Walkthrough:

(a) Flip n coins and let $X_i = x_i$ if the i^{th} coin comes up heads, and $X_i = 0$ if it comes up tails. Let A be the set of indices of coins that came up heads. Write

$$\sum_{i=1}^{n} X_i = S_A.$$

(b) What is $\mathrm{E}[X_i]$? $\mathrm{E}[S_A]$? $\mathrm{Var}(X_i)$? $\mathrm{Var}(S_A)$?

(c) What does Chebyshev's inequality (**11**) say when you plug in 2λ?

(d) Show that $P(S_A \geq \lambda) = P(-S_A \geq \lambda)$.

(e) Conclude that the inequality given in the problem holds.

(f) What does the equality case of Chebyshev's inequality say about the equality case for this problem? Conclude.

Day 2: Analytic Number Theory

Definition 13. For functions f and g where there exists constants N and C such that $|g(n)| \leq Cf(n)$ for all $n > N$, we can write $g = O(f)$. In other words, $g = O(f)$ if and only if $f(n)$ is positive for sufficiently large n and $\limsup_{n\to\infty} \left|\frac{g(n)}{f(n)}\right| < \infty$.

Definition 14. For functions f and g, if for every $\epsilon > 0$ there exists some N such that $|g(n)| \leq \epsilon f(n)$ for all $n > N$, we can write $g = o(f)$. In other words, $g = o(f)$ if and only if $f(n)$ is positive for sufficiently large n and $\lim_{n\to\infty} \left|\frac{g(n)}{f(n)}\right| = 0$.

Definition 15. For functions f and g, if for every $\epsilon > 0$ there exists some N such that $\max\left(\frac{f(n)}{g(n)}, \frac{g(n)}{f(n)}\right) < 1+\epsilon$, we can write $g \sim f$. In other words, $g \sim f$. if and only if $\lim\limits_{n\to\infty} \frac{g(n)}{f(n)} = 1$.

(If you don't understand the limit definitions, don't worry about it.)

Definition 16. The *prime counting function* $\pi(n)$ is the number of primes less than or equal to n.

The prime number theorem states that $\pi(n) \sim \frac{n}{\ln n}$. We prove the following weaker statement.

Theorem 17 (Chebyshev)**.**

$$\pi(n) = O\Big(\frac{n}{\ln n}\Big)$$

Walkthrough: (Variant of Erdős).

(a) Show that if $n < p \leq 2n$, $p \,\big|\, \binom{2n}{n}$. Conclude that $n^{\pi(2n)-\pi(n)} < 4^n$, and that $\pi\big(2^k\big) - \pi\big(2^{k-1}\big) < \frac{2^k}{k-1}$.

(b) Show that $\pi(4^m) - 1 < \sum\limits_{k=2}^{2m} \frac{2^k}{k-1}$, and conclude that $\pi(4^m) < 2^{m+1} + \frac{2^{2m+1}}{m}$.

(c) Conclude.

Definition 18. The *von Mangoldt function* is

$$\Lambda(n) := \begin{cases} \ln p & \text{if } n > 1 \text{ is a power of } p \\ 0 & \text{otherwise.} \end{cases}$$

Homework 19. Show the following key property of this function:

$$\sum_{d|n} \Lambda(d) = \ln n.$$

Definition 20. The *second Chebyshev function* is

$$\psi(x) := \sum_{n\leq x} \Lambda(n).$$

The prime number theorem is equivalent to $\psi(x) \sim x$. We prove the following weaker statement.

Theorem 21.

$$\psi(x) = O(x).$$

Proof. By Chebyshev (**17**),

$$\psi(x) = \sum_{n\le x} \Lambda(n) = \sum_{p\le x}\sum_{p^a\le x} \ln p = \sum_p \ln\left(p^{\lfloor \log_p(x)\rfloor}\right) \le \sum_p \ln(x) = (\ln x)\sum_p 1 = (\ln x)\pi(x) = O(x).$$

□

We now introduce a very useful tool in analytic number theory. Be warned that the following discussion does involve some calculus.

Theorem 22 (Abel summation formula)**.** *Given* $(a_n)_{n=1}^{\infty}$*, let*

$$A(x) = \sum_{n\le x} a_n.$$

If f *is a continuous function for* $x \ge 1$*,*

$$\sum_{n\le x} a_n f(n) = A(x)f(x) - \int_1^x A(t)f'(t)\,\mathrm{d}t.$$

You can think of this as a more discrete version of integration by parts.

Bonus Homework 23. Prove this.

Theorem 24 (Weak version of Stirling's formula)**.**

$$\ln x! = x\ln x - x + O(\ln x)$$

Proof. Using $a_n = 1$, $A(x) = \lfloor x\rfloor$ and $f(x) = \ln(x)$ in the Abel Summation formula (**22**),

$$\begin{aligned}\ln x! = \sum_{n\le x}\ln n &= x\ln x - \int_1^x \frac{\lfloor x\rfloor}{x} = x\ln x - \int_1^x \frac{x-\{x\}}{x}\\ &= x\ln x - \int_1^x 1 + \int_1^x \frac{\{x\}}{x} = x\ln x - (x-1) + O\left(\int_1^x \frac{1}{x}\right)\\ &= x\ln x - x + O(\ln x)\end{aligned}$$

□

Theorem 25.

$$\sum_{d\le x}\frac{\Lambda(d)}{d} = \ln x + O(1).$$

Proof. By homework **19** and theorems **21** and **24**,

$$x\ln x + o(x\ln x) = \sum_{n\le x}\ln n = \sum_{n\le x}\sum_{d\mid n}\Lambda(d) = \sum_{d\le x}\left\lfloor\frac{x}{d}\right\rfloor\Lambda(d) = \sum_{d\le x} x\frac{\Lambda(d)}{d} + O(\psi(x)) = x\sum_{d\le x}\frac{\Lambda(d)}{d} + O(x).$$

Dividing by x, we get the desired result. □

Homework 26. Show that

$$\sum_{n\in\mathbb{N}} \frac{\ln n}{n^2}$$

converges.

Theorem 27 (Merten's first theorem)**.**

$$\sum_{p\le x} \frac{\ln p}{p} = \ln x + O(1).$$

Proof. By Theorem **25** and Homework **26**,

$$\begin{aligned}\ln x + O(1) &= \sum_{d\le x} \frac{\Lambda(d)}{d} = \sum_{p\le x}\sum_{p^a\le x} \frac{\ln p}{p^a} \\ &= \sum_{p\le x} \frac{\ln p}{p} + O\left(\sum_{p\le x} (\ln p)\left(\frac{1}{p^2}+\frac{1}{p^3}+\cdots\right)\right) \\ &= \sum_{p\le x} \frac{\ln p}{p} + O\left(\sum_{p\le x} \frac{2\ln p}{p^2}\right) = \sum_{p\le x} \frac{\ln p}{p} + O(1),\end{aligned}$$

giving the desired result. □

Theorem 28 (Merten's second theorem)**.** *There exists a constant M such that*

$$\sum_{p\le x} \frac{1}{p} = \ln\ln x + M + O\left(\frac{1}{\ln x}\right).$$

Proof. Using $a_p = \frac{\ln p}{p}$ and $a_n = 0$ for n not prime, and $f(x) = \frac{1}{\ln x}$, in the Abel summation formula (**22**), we have that by Merten's first theorem (**27**),

$$\begin{aligned}\sum_{p\le x} \frac{1}{p} &= \frac{A(x)}{\ln x} + \int_1^x \frac{A(t)}{t\ln^2 t}\,\mathrm{d}t = \frac{A(x)}{\ln x} + \int_2^x \frac{A(t)}{t\ln^2 t}\,\mathrm{d}t \\ &= \frac{\ln x + O(1)}{\ln x} + \int_2^x \frac{\ln t + O(1)}{t\ln^2 t}\,\mathrm{d}t \\ &= \left(1 + O\left(\frac{1}{\ln x}\right)\right) + \int_2^x \frac{1}{t\ln t}\,\mathrm{d}t + \int_2^x \frac{O(1)}{t\ln^2 t}\,\mathrm{d}t \\ &= \left(1 + O\left(\frac{1}{\ln x}\right)\right) + (\ln\ln x - \ln\ln 2) + \left(\int_2^\infty \frac{O(1)}{t\ln^2 t}\,\mathrm{d}t - \int_x^\infty \frac{O(1)}{t\ln^2 t}\,\mathrm{d}t\right) \\ &= \ln\ln x + \left(1 - \ln\ln 2 + \int_2^\infty \frac{O(1)}{t\ln^2 t}\,\mathrm{d}t\right) + O\left(\frac{1}{\ln x}\right),\end{aligned}$$

as desired. □

M is known as the *Meissel-Mertens constant*, and has value approximately 0.2615.

Day 3: Putting it all together

We turn our attention to various facts about the sums of random variables. The most fundamental such statement is linearity of expectation, namely, that the expected value of a sum of random variables is the sum of the expected values of the random variables.

The variance of a sum is as follows:

Homework 29. Show that for random variables X_i,

$$\operatorname{Var}\left(\sum X_i\right) = \sum \operatorname{Var}(X_i) + \sum_{i<j} 2\operatorname{Cov}(X_i, X_j).$$

One of the most important results about the sums of random variables is the Central Limit Theorem, which we will not prove, as it requires learning about moment-generating functions. (I do encourage you to read about them on your own time.) The theorem roughly states that the sum of many independent random variables can be approximated by a normal distribution, as long as the variables do not get too extreme. One version of the CLT is precisely stated as follows:

Theorem 30 (Lyapunov Central Limit Theorem)**.** *Suppose $\{X_1, \ldots, X_n\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance V_i. Define $s_n = \sqrt{\sum_{i=1}^{n} V_i}$.*

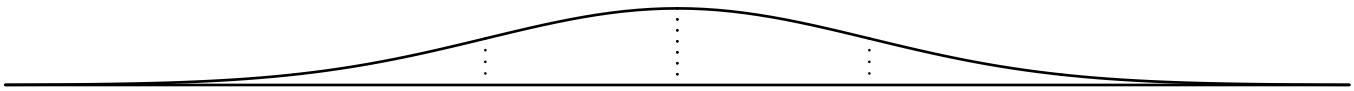
If for some $\delta > 0$, Lyapunov's condition

$$\sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^{2+\delta}\right] = o\left(s_n^{2+\delta}\right)$$

is satisfied, then a sum of $\frac{X_i - \mu_i}{s_n}$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{d} N(0,1).$$

Here is an illustration of the normal distribution:



We now state the fundamental theorem of probabilistic number theory.

Theorem 31 (Erdős - Kac)**.** *Let $\nu(n)$ be the number of distinct prime divisors of n, and let K_n be the number of integers m from 1 to n that satisfy $\nu(m) < \ln\ln + \omega\sqrt{\ln\ln n}$. Then,*

$$\lim_{n\to\infty} \frac{K_n}{n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\omega} e^{-t^2/2}\,\mathrm{d}t.$$

This theorem basically says $\nu(n)$ is normally distributed with expected value $\ln\ln n$ and standard deviation $\sqrt{\ln\ln n}$. In particular, for $n = 10^{10^{44}}$, $\ln\ln n \approx 102$, and numerically, the distribution of $\nu(m)$ has average and standard deviation about 100 and 10, respectively.

While we will not cover everything necessary for the proof of this statement, we will fully prove the following earlier and weaker result found by Hardy and Ramanujan in 1917.

Theorem 32 (Hardy-Ramanujan)**.** *Let $\omega(n) \to \infty$ arbitrarily slowly. Then the number of x in $\{1, \ldots, n\}$ such that*

$$|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}$$

is $o(n)$.

Here is a proof of the Hardy-Ramanujan theorem found by Turán in 1934:

Proof. Let $D = n^{1/137}$, and let $\nu'(m)$ be the number of prime factors of m that are $\leq D$. We have $\nu(m) > \nu'(m) > \nu(m) - 137$, so as $137 = o\left(\sqrt{\ln\ln n}\right)$, it suffices to show that the statement of this theorem is true for ν'.

Choose a random number m from 1 to n. Let X_p be 1 if $p \,\big|\, m$ and 0 otherwise. We write

$$\nu'(m) = \sum_{p \leq D} X_p.$$

We have

$$\mathrm{E}\left[X_p\right] = \frac{\lfloor n/p \rfloor}{n} = \frac{n/p - \{n/p\}}{n} = 1/p + O(1/n),$$

so by Merten's second theorem (**28**),

$$\mathrm{E}\left[\nu'(m)\right] = \sum_{p \leq D}\left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \ln\ln n + O(1) + O\left(\frac{D}{n}\right) = \ln\ln n + O(1).$$

By Homework **29**, we also have

$$\begin{aligned}
\mathrm{Var}\left(\nu'(m)\right) &= \sum_{p \leq D} \mathrm{Var}\left(X_p\right) + \sum_{p<q\leq D} 2\,\mathrm{Cov}\left(X_p, X_q\right) \\
&= \sum_{p \leq D} \frac{\lfloor n/p \rfloor}{n}\left(1 - \frac{\lfloor n/p \rfloor}{n}\right) + \sum_{p<q\leq D} 2(\mathrm{E}\left[X_p, X_q\right] - \mathrm{E}\left[X_p\right]\mathrm{E}\left[X_q\right]) \\
&\leq \sum_{p \leq D} \frac{1}{p} + \sum_{p<q\leq D} 2\left(\frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n}\right) \\
&\leq \ln\ln n + O(1) + \sum_{p<q\leq D} 2\left(\frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right)\right) \\
&\leq \ln\ln n + O(1) + \sum_{p<q\leq D} \frac{2}{n}\left(\frac{1}{p} + \frac{1}{q}\right) \\
&\leq \ln\ln n + O(1) + \frac{2D}{n}\sum_{p \leq D}\frac{1}{p} \\
&= \ln\ln n + O(1) + O\left(\frac{D}{n}(\ln\ln n + O(1))\right) \\
&= \ln\ln n + O(1),
\end{aligned}$$

again, by Merten's second theorem. Thus, by Chebyshev's inequality (**11**),

$$P\left(|\nu(x) - \ln\ln x| > \omega(n)\sqrt{\ln\ln n}\right) < \frac{1}{\omega(n)^2} + o(1) = o(1),$$

as desired. □

Here is the proof of the full-fledged Erdős-Kac theorem.

- The n^{th} moment of a random variable X is $\mathrm{E}[X^n]$. For example, the first moment of a variable X is expected value, the centralized second moment variance, the standardized third moment skewness, and the standardized fourth moment kurtosis.

  It is known that if the moments of a sequence of distributions converge to the moments of another distribution, its cumulative distribution is the limiting cumulative distribution. (We will also not show this, as it again requires moment-generating functions.) Thus, it suffices to show that the moments of ν when normalized match the moments of the standard normal distribution.
- Set D to be $n^{1/\ln\ln\ln n}$, as $\ln\ln\ln n$ is unbounded but grows slower than $\sqrt{\ln\ln n}$, and define ν' as above. The reason for it to grow slower than $\sqrt{\ln\ln n}$ is to not affect the true value of ν too much, and the reason to grow unbounded is so that the various moments will not be affected by the finite sample size.
- For $p < D$, let Y_p be 1 with probability $1/p$ and 0 otherwise, sort of an idealized version of X_p. Define $\nu_0 := \sum_{p \le D} Y_p$. Set

  $$\mu := \mathrm{E}[Y] \approx \sum_{p \le D} \frac{1}{p} = \ln\ln n + O(\ln\ln\ln\ln n) = \ln\ln n + o\Big((\ln\ln n)^{1/2}\Big)$$

  and

  $$\sigma^2 := \mathrm{Var}(Y) = \sum_{p \le D} \frac{1}{p}\left(1 - \frac{1}{p}\right) = \ln\ln n + O(\ln\ln\ln\ln n) \sim \ln\ln n,$$

  and normalize to get $\widetilde{\nu'} := \frac{\nu' - \mu}{\sigma}$ and $\widetilde{\nu_0} := \frac{\nu_0 - \mu}{\sigma}$.

  Homework 33. Check that Lyapunov's condition does indeed hold on the Y_i's.

  By the Lyapunov Central Limit Theorem (**30**), $\nu_0 \xrightarrow{d} N$, where N is the standard normal distribution, so $\mathrm{E}\left[\widetilde{\nu_0}^k\right] \to \mathrm{E}\left[N^k\right]$ for every positive integer k. We now compare $\mathrm{E}\left[\widetilde{\nu'}^k\right]$ and $\mathrm{E}\left[\widetilde{\nu_0}^k\right]$.
- $\left(\widetilde{\nu'}\right)^k = \left(\frac{\nu' - \mu}{\sigma}\right)^k$ is a polynomial in ν' with coefficients of size $O\Big((\ln\ln n)^{k/2}\Big) = n^{o(1)}$. Furthermore, writing ν' as $\sum_{p \le D} X_p$ and expanding gives $O\Big(\pi(D)^k\Big) = O\big(D^k\big) = n^{o(1)}$ terms of the form $X_{p_1} \cdots X_{p_s}$ because $X_p^j = X_p$ for all p, j.
- We have for all distinct primes $p_1, \ldots, p_k < D$,

  $$\mathrm{E}[X_{p_1} \cdots X_{p_k}] - \mathrm{E}[Y_{p_1} \cdots Y_{p_k}] = \frac{1}{n}\left(\left\lfloor \frac{n}{p_1 \cdots p_k} \right\rfloor - \frac{n}{p_1 \cdots p_k}\right) = O\left(\frac{1}{n}\right),$$

  so

  $$\mathrm{E}\left[\widetilde{\nu'}^k\right] - \mathrm{E}\left[\widetilde{\nu_0}^k\right] = n^{o(1)} n^{o(1)} O\left(\frac{1}{n}\right) = o(1),$$

  and the moments of $\widetilde{\nu'}$ and $\widetilde{\nu_0}$ are indeed the same.

To summarize this proof, we took the primes less than $D = n^{1/\ln\ln\ln n}$ and modeled their corresponding X_p's with Y_p's. By a variant of the central limit theorem, the sum of these Y_p's approach the normal distribution for large n. In addition, any k of these X_p's are independent for large n, as $\ln\ln\ln n \to \infty$, so the Y_p's are a good approximation of the X_p's. Together, these produce the Erdős-Kac theorem, a beautiful interaction between probability and number theory.

Bonus Homework 34. Would the Erdős-Kac theorem still hold if $\nu(n)$ instead counted the number of prime divisors of n *with multiplicity*?