# AN INVITATION TO MODERN NUMBER THEORY

STEVEN J. MILLER AND
RAMIN TAKLOO-BIGHASH

ABSTRACT. These notes are a corrected proof of the Central Limit Theorem for independent identically distributed binomial random variables with probability 1/2 of success (Chapter 8, pages 214–215 of the first edition).

## 1. CHEBYSHEV'S INEQUALITY

**Exercise 1.1** (Chebyshev's Inequality)**.** *Let $X$ be a random variable with mean $\mu$ and finite variance $\sigma^2$. Prove Chebyshev's inequality:*

$$\text{Prob}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \tag{1}$$

*where $\text{Prob}(|X - \mu| \geq a)$ is the probability that $X$ takes on values at least $a$ units from the mean. Chebyshev's theorem holds for all nice distributions, and provides bounds for being far away from the mean (where far is relative to the natural spacing, namely $\sigma$).*

## 2. PROOF FOR BERNOULLI PROCESSES

We sketch the proof of the Central Limit Theorem for Bernoulli Processes where the probability of success is $p = \frac{1}{2}$. Consider the random variable $X$ that is 1 with probability $\frac{1}{2}$ and $-1$ with probability $\frac{1}{2}$ (for example, tosses of a fair coin; the advantage of making a tail $-1$ is that the mean is zero). Note the mean of $X$ is $\overline{X} = 0$, the variance is $\sigma_X^2 = 1$ (as we have $1^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2}$) and the standard deviation is $\sigma_X = 1$.

Let $X_1, \ldots, X_{2N}$ be independent identically distributed random variables, distributed as $X$ (it simplifies the expressions to consider an even number of tosses). Consider $S_{2N} = X_1 + \cdots + X_{2N}$. Its mean is zero and its variance is $2N$, and we expect fluctuations of size $\sqrt{2N}$. We show that for $N$ large the distribution of $S_{2N}$ is approximately normal. We need

**Lemma 2.1** (Stirling's Formula)**.** *For $n$ large,*

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + O(1/n)\right). \tag{2}$$

For a proof, see [WW]. We show (2) is a reasonable approximation. It is often easier to analyze a product by converting it to a sum; this is readily accomplished by taking logarithms. We have

$$\log n! \;=\; \sum_{k=1}^{n} \log k \;\approx\; \int_1^n \log t dt \;=\; (t\log t - t)|_1^n. \tag{3}$$

Thus $\log n! \approx n\log n - n$, or $n! \approx n^n e^{-n}$.

We now consider the distribution of $S_{2N}$. We first note that the probability that $S_{2N} = 2k+1$ is zero. This is because $S_{2N}$ equals the number of heads minus the number of tails, which is always even: if we have $k$ heads and $2N-k$ tails then $S_{2N}$ equals $2N-2k$.

The probability that $S_{2N}$ equals $2k$ is just $\binom{2N}{N+k}(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$. This is because for $S_{2N}$ to equal $2k$, we need $2k$ more 1's (heads) than $-1$'s (tails), and the number of 1's and $-1$'s add to $2N$. Thus we have $N+k$ heads (1's) and $N-k$ tails ($-1$'s). There are $2^{2N}$ strings of 1's and $-1$'s, $\binom{2N}{N+k}$ have exactly $N+k$ heads and $N-k$ tails, and the probability of each string is $(\frac{1}{2})^{2N}$. We have written $(\frac{1}{2})^{N+k}(\frac{1}{2})^{N-k}$ to show how to handle the more general case when there is a probability $p$ of heads and $1-p$ of tails.

We use Stirling's Formula to approximate $\binom{2N}{N+k}$. After elementary algebra we find

$$\begin{aligned}\binom{2N}{N+k} &\approx \frac{(2N)^{2N}}{(N+k)^{N+k}(N-k)^{N-k}}\sqrt{\frac{N}{\pi(N+k)(N-k)}} \\ &= \frac{2^{2N}}{\sqrt{\pi N}}\frac{1}{(1+\frac{k}{N})^{N+\frac{1}{2}+k}(1-\frac{k}{N})^{N+\frac{1}{2}-k}}.\end{aligned} \tag{4}$$

We would like to use $\left(1+\frac{w}{N}\right)^N \approx e^w$ from §5.4; unfortunately, we must be a little more careful as the values of $k$ we consider grow with $N$. For example, we might believe that $(1+\frac{k}{N})^N \to e^k$ and $(1-\frac{k}{N})^N \to e^{-k}$, so these factors cancel. As $k$ is small relative to $N$ we may ignore the factors of $\frac{1}{2}$, and then say

$$\left(1+\frac{k}{N}\right)^k \;=\; \left(1+\frac{k}{N}\right)^{N\cdot\frac{k}{N}} \;\to\; e^{k^2/N}; \tag{5}$$

similarly, $(1-\frac{k}{N})^{-k} \to e^{k^2/N}$. Thus we would claim (*and we shall see later in Lemma 2.2 that this claim is in error!*) that

$$\left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \;\to\; e^{2k^2/N}. \tag{6}$$

We show that $\left(1+\frac{k}{N}\right)^{N+\frac{1}{2}+k}\left(1-\frac{k}{N}\right)^{N+\frac{1}{2}-k} \to e^{k^2/N}$. The importance of this calculation is that it highlights how crucial rates of convergence are.

While it is true that the main terms of $(1 \pm \frac{k}{N})^N$ are $e^{\pm k}$, the error terms (in the convergence) are quite important, and yield large secondary terms when $k$ is a power of $N$. What happens here is that the secondary terms from these two factors reinforce each other. Instead of using $\left(1 + \frac{w}{N}\right)^N \approx e^w$ from §5.4, it is better to take the logarithms of the two factors, Taylor expand, and then exponentiate. This allows us to better keep track of the error terms.

An immediate consequence of Chebyshev's inequality (see Exercise 1.1) is that we need only study $k$ where $|k|$ is at most $N^{\frac{1}{2}+\epsilon}$. This is because the standard deviation of $S_{2N}$ is $\sqrt{2N}$. Specifically, see Exercise 2.4 for a proof that given any $\epsilon > 0$, the probability of observing a $k$ with $|k| \gg N^{\frac{1}{2}+\epsilon}$ is negligible. Thus it suffices to analyze the probability that $S_{2N} = 2k$ for $|k| \leq N^{\frac{1}{2}+\frac{1}{9}}$.

**Lemma 2.2.** *For any $\epsilon \leq \frac{1}{9}$, for $N \to \infty$ with $k \ll N^{\frac{1}{2}+\epsilon}$, we have*

$$\left(1 + \frac{k}{N}\right)^{N+\frac{1}{2}+k} \left(1 - \frac{k}{N}\right)^{N+\frac{1}{2}-k} \to e^{k^2/N} e^{O(N^{-1/6})}. \tag{7}$$

*Proof.* Recall that for $|x| < 1$,

$$\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^n}{n}. \tag{8}$$

As we are assuming $k \ll N^{\frac{1}{2}+\epsilon}$, note that any term below of size $k^2/N^2$, $k^3/N^2$ or $k^4/N^3$ will be negligible. Thus we have

$$\begin{aligned}
P_{k,N} &= \left(1 + \frac{k}{N}\right)^{N+\frac{1}{2}+k} \left(1 - \frac{k}{N}\right)^{N+\frac{1}{2}-k} \\
\log P_{k,N} &= \left(N + \frac{1}{2} + k\right) \log\left(1 + \frac{k}{N}\right) + \left(N + \frac{1}{2} - k\right) \log\left(1 - \frac{k}{N}\right)^{N+\frac{1}{2}-k} \\
&= \left(N + \frac{1}{2} + k\right)\left(\frac{k}{N} - \frac{k^2}{2N^2} + O\left(\frac{k^3}{N^3}\right)\right) \\
&\quad + \left(N + \frac{1}{2} - k\right)\left(-\frac{k}{N} - \frac{k^2}{2N^2} + O\left(\frac{k^3}{N^3}\right)\right) \\
&= \frac{2k^2}{N} - 2\left(N + \frac{1}{2}\right)\frac{k^2}{2N^2} + O\left(\frac{k^3}{N^2} + \frac{k^4}{N^3}\right) \\
&= \frac{k^2}{N} + O\left(\frac{k^2}{N^2} + \frac{k^3}{N^2} + \frac{k^4}{N^3}\right).
\end{aligned} \tag{9}$$

As $k \ll N^{\frac{1}{2}+\epsilon}$, for $\epsilon < \frac{1}{9}$ the big-Oh term is dominated by $N^{-1/6}$, and we finally obtain that

$$P_{k,N} \quad = \quad e^{k^2/N} e^{O\left(N^{-1/6}\right)}, \tag{10}$$

which completes the proof. □

Combining Lemma 2.2 with (4) yields

$$\binom{2N}{N+k}\frac{1}{2^{2N}} \quad \approx \quad \frac{1}{\sqrt{\pi N}}\, e^{-k^2/N}. \tag{11}$$

The proof of the central limit theorem in this case is completed by some simple algebra. We are studying $S_{2N} = 2k$, so we should replace $k^2$ with $(2k)^2/4$. Similarly, since the variance of $S_{2N}$ is $2N$, we should replace $N$ with $(2N)/2$. We find

$$\text{Prob}(S_{2N} = 2k) \;=\; \binom{2N}{N+k}\frac{1}{2^{2N}} \quad \approx \quad \frac{2}{\sqrt{2\pi \cdot (2N)}}\, e^{-(2k)^2/2(2N)} \tag{12}$$

Remember $S_{2N}$ is never odd. The factor of 2 in the numerator of the normalization constant above reflects this fact, namely the contribution from the probability that $S_{2N}$ is even is twice as large as we would expect, because it has to account for the fact that the probability that $S_{2N}$ is odd is zero. Thus the above looks like a Gaussian with mean 0 and variance $2N$. For $N$ large such a Gaussian is slowly varying, and integrating from $2k$ to $2k+2$ is basically $2/\sqrt{2\pi(2N)} \;\cdot\; \exp -(2k)^2/2(2N)$.

**Exercise 2.3.** *Use the integral test to bound the error in* (3)*, and then use that to bound the error in the estimate of* $n!$.

**Exercise 2.4.** *Prove the standard deviation of* $S_{2N}$ *is* $\sqrt{2N}$*. Use this and Chebyshev's inequality (Exercise 1.1) to prove*

$$\text{Prob}(|S_{2N}| \geq N^{\epsilon} \cdot \sqrt{2N}) \;\leq\; \frac{1}{N^{2\epsilon}}, \tag{13}$$

*which implies that it suffices to study values of* $k$ *with* $k \ll N^{\frac{1}{2}+\epsilon}$.

**Exercise 2.5.** *Prove* (8).

**Exercise 2.6.** *Can you generalize the above arguments to handle the case when* $p \neq \frac{1}{2}$.

## REFERENCES

[WW] E. Whittaker and G. Watson, *A Course of Modern Analysis*, 4th edition, Cambridge University Press, Cambridge, 1996.