

A Simple and Combinatorial Approach to Proving Chernoff Bounds and Their Generalizations

William Kuszmaul*

Abstract

The Chernoff bound is one of the most widely used tools in theoretical computer science. It's rare to find a randomized algorithm that *doesn't* employ a Chernoff bound in its analysis.

The standard proofs of Chernoff bounds are beautiful but in some ways not very intuitive. In this paper, I'll show you a different proof that has four features:

- the proof offers a strong intuition for why Chernoff bounds look the way that they do;
- the proof is user-friendly and (almost) algebra-free;
- the proof comes with matching lower bounds, up to constant factors in the exponent;
- the proof extends to establish generalizations of Chernoff bounds in other settings.

The ultimate goal is that, once you know this proof (and with a bit of practice), you should be able to confidently reason about Chernoff-style bounds in your head, extending them to other settings, and convincing yourself that the bounds you're obtaining are tight (up to constant factors in the exponent).

1 Introduction

The Chernoff bound is one of the most widely used tools in theoretical computer science. It's rare to find a randomized algorithm that *doesn't* employ a Chernoff bound in its analysis.

The standard proofs of Chernoff bounds are beautiful but in some ways not very intuitive. In this paper, I'll show you a different proof that, as far as I can tell, has not appeared in the literature. The proof will have four noteworthy features:

1. **Intuition:** The proof offers a clean intuition for why Chernoff bounds have the shapes that they do.
2. **Lower Bounds:** The proof comes with matching *lower bounds*. In fact, one way to perform the proof is to first prove lower bounds, and then directly argue that those lower bounds are tight (up to constants in the exponent).
3. **User-Friendliness:** Most proofs of Chernoff bounds require various identities (and Taylor series approximations) to obtain the final user-friendly bound. The proof here will lead directly to a user-friendly bound without requiring intermediate algebra.
4. **Generality:** The proof can be extended to establish many of the classical generalizations of Chernoff bounds (e.g., Hoeffding bounds, Azuma's inequality, Bernstein-type inequalities, Bennett's inequality, etc.). In other words, the proof isn't just a party trick.

The proof will also have a noteworthy drawback: It will give the correct bound *up to constant factors in the exponent*. If those constant factors are important to you, then you'll want to use other derivations.

How to think about Chernoff bounds. Before getting into proofs, it is worth first reviewing the basic bounds that we will wish to prove. Let X_1, X_2, \dots, X_n be independent 0-1 random variables satisfying $\Pr[X_i = 1] = p$ for some $p \leq 1/2$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X] = np$.

The Chernoff bound tells us that the probability of X deviating substantially above its mean μ is small. That is, we get an upper bound on $\Pr[X \geq \mu + t]$ as a function of t .

*Carnegie Mellon University.

Chernoff bounds are presented in many different forms, and students often have trouble figuring out which version to memorize (the result is that many students end up using Wikipedia as a regular reference). So what is the right way to think about Chernoff bounds?

The first thing to know is that n and p are red herrings. The only parameter that actually matters is μ . In fact, one can obtain tight Chernoff bounds in all regimes by just remembering two simple bounds. The first is the **small-deviation bound**, which says that

$$(1.1) \quad \Pr[X \geq \mu + k\sqrt{\mu}] = \frac{1}{2^{\Theta(k^2)}}$$

for any $k = O(\sqrt{\mu})$ satisfying $k\sqrt{\mu} \leq n$. The second is the **large-deviation bound**, which says that

$$(1.2) \quad \Pr[X \geq \mu + r\mu] = \frac{1}{\Theta(r)^{\Theta(r\mu)}}$$

for any $r \geq 1$ satisfying $\mu + r\mu \leq n$.

There are three things to notice about these bounds. First, as I mentioned earlier, we are not worrying about what the constants are in the exponents. As theoreticians, we are almost always interested in asymptotic deviations (i.e., $\Pr[X \geq \mu + \Omega(t)]$ for various t), so the constant in the exponent typically doesn't matter. Second, the fact that the exponents are in Θ -notation, rather than Ω -notation, is no coincidence: both bounds turn out to be tight up to constant factors in the exponent. Third, the two bounds become equivalent when we consider $\Pr[X \geq \mu + \Theta(\mu)]$, so there is a smooth transition from one regime to the other.

It is also worth taking a few moments to internalize the shapes of these bounds. The small-deviation bound, should be viewed as telling us something about standard deviations. It turns out that the standard deviation of X is guaranteed to be $\Theta(\sqrt{\mu})$, regardless of n and p (as long as $p \leq 1/2$). Thus the bound says that the probability of being k standard deviations above the mean shrinks at a rate of $1/2^{\Theta(k^2)}$. In fact, we will later see that the previous sentence continues to be true in much more general settings, and that this is the source of what is known as Bennett's inequality (we will come back to this later).

The large-deviation bound also takes an interesting shape. It is much stronger than most students would guess it should be. A priori, students typically assume that the bound should be something like $1/2^{\Theta(r\mu)}$. This is correct when $r = \Theta(1)$, but when r is larger, (1.2) gets stronger, replacing the denominator of 2 with r .

Although the above Chernoff bounds are stated in the case where X_1, X_2, \dots, X_n are identically distributed 0-1 random variables, the same upper bounds hold for any independent real-valued random variables $X_1, X_2, \dots, X_n \in [0, 1]$ satisfying $\mathbb{E}[\sum_i X_i] = \mu$. The corresponding lower bounds do not necessarily hold in this more general setting (for example, it might be that X_1, X_2, \dots, X_n are all deterministically 1, so $\Pr[X = \mu] = 1$), but we shall see that it is a relatively simple task to reason about when the lower bounds do or do not hold.

Paper outline. In the body of the paper, we will present the new Chernoff bound derivation from four different perspectives:

- **The One-Page Version (Section 2).** We begin in Section 2 with a bare-bones version of the proof—a one-page self-contained analysis that focuses on the special case where we have n fair coin flips. This version of the proof is designed for readers who like to read first and digest after. It does not concern itself with side-quests such as proving lower bounds or highlighting intuition. Additionally, to simplify the presentation, and because we are focusing only on *fair* coin flips, we follow the convention in both this section and the next that each X_i is in $\{-1, 1\}$ rather than $\{0, 1\}$.
- **The Extended Edition (Section 3).** In Section 3, we present the same proof again, but with additional commentary to motivate the steps and explain what's going on at a higher level. This version of the proof is designed for readers who like digest as they read. It includes a focus on intuition, as well as a small side-quest to prove matching lower bounds. In fact, quite happily, the lower-bound proof serves as a strong motivator for why the *upper-bound proof* should follow the structure that it does.
- **Bias Coin Flips and the Large-Deviation Regime (Section 4).** Section 4 extends the proof to the setting of biased coin flips, where each coin has some probability $p \leq 1/2$ of being heads. This allows us to present the large-deviation bound (Equation 1.2). The proof follows a very similar structure to the small-deviation case, and comes once again with matching lower bounds.

- **Generalizing to an Adaptive Bennett’s Inequality (Section 5).** Having proven both the small and large deviation bounds for classical Chernoff bounds, we turn our attention in Section 5 to proving a powerful generalization of Chernoff bounds, namely, an adaptive version of Bennett’s Inequality. Here, we are intentionally picking one of the most “heavy-weight” generalizations of Chernoff bounds. The point is to demonstrate how, with the same basic techniques that we used to prove the basic Chernoff bounds, and by just filling in a few more details, we can walk away with bounds that would traditionally be viewed as out of reach for combinatorial proofs.

We remark that the one-page proof in Section 2 is short but is not necessarily the right starting place for every reader. Some readers (especially students seeing Chernoff bounds for the first time) may wish to start with Sections 3 and 4, and then to optionally add on additional sections from there.

The ultimate goal of the paper is that, once you have digested the proof (and with a bit of practice), you should be able to confidently reason about Chernoff-style bounds in your head, extending them to other settings, and convincing yourself that the bounds you’re obtaining are tight (up to constant factors in the exponent).

Historical context and past work. Chernoff bounds first appeared in the literature in 1952 paper by Herman Chernoff [6] (although Chernoff himself attributes them to Herman Rubin [7]). The bounds and their generalizations have also been independently formulated by many other authors, including Kazuoki Azuma [1], Wassily Hoeffding [15], and Sergei Bernstein [4].

The classical proof of Chernoff bounds proceeds by applying Markov’s inequality to the moment-generating function of a random variable. This is an important technique, that has also served as a core foundation for much of the work on concentration inequalities in statistics and probability theory [2, 10–14, 17, 19–22, 25–27, 29] (see [5] or [8] for a survey). There have also been several other proofs [9, 16, 23, 28], using techniques from areas ranging from coding theory [23] to differential privacy [28]; see Mulzer’s survey [24] for a description of the five main proof approaches that have been proposed.

Most of these proof approaches [24] struggle to generalize to more diverse settings—indeed, besides the classical moment-generating function approach, only one of the other approaches covered in [24], namely the proof of [16], appears to extend to prove Azuma’s inequality (which, in turn, is weaker than the generalization that we prove in Section 5). Additionally, all of the previous proofs [24] share the unfortunate property that, in order to get to a user-friendly bound (i.e., to either of Equations (1.1) or (1.2)), one must first apply algebraic identities such as Taylor expansions.

There is, not surprisingly, much less of a focus on lower bounds than there are on upper bounds. The classical moment-generating-function argument can be extended (non-trivially) to obtain essentially matching lower bounds see, e.g., [18, 30]. The simple combinatorial approach to proving lower bounds that is taken in the current paper does not appear to have been observed in past work, and the most general lower bound that we prove (Section 5.4) does not appear to follow from the standard lower-bound techniques [18, 30].

2 Fair Coin Flips: The Bare-Bones Proof

In this section, we consider a sum $X = \sum_{i=1}^n X_i$ of independent unbiased coin flips $X_i \in \{1, -1\}$, and we prove that $\Pr[X \geq k\sqrt{n}] \leq 2^{-\Omega(k^2)}$. Our starting point is a simple extension of Chebyshev's inequality:

LEMMA 2.1. (EXTENDED CHEBYSHEV) *For any $k \geq 1$, we have $\Pr[\max_j \sum_{i=1}^j X_i \geq k\sqrt{n}] \leq \frac{2}{k^2}$.*

Proof. By Chebyshev's inequality, we have $\Pr[\sum_{i=1}^n X_i \geq k\sqrt{n}] \leq 1/k^2$. Thus, it suffices to show that

$$(2.3) \quad \Pr \left[\max_j \sum_{i=1}^j X_i \geq k\sqrt{n} \right] \leq 2 \Pr \left[\sum_{i=1}^n X_i \geq k\sqrt{n} \right].$$

On the other hand, (2.3) follows from the following simple observation: If there exists $j \geq 0$ such that $\sum_{i=1}^j X_i \geq k\sqrt{n}$, then with probability at least 0.5 we have that $\sum_{i=j+1}^n X_i \geq 0$, and thus that $\sum_{i=1}^n X_i \geq k\sqrt{n}$. \square

Using Lemma 2.1, we can derive a very simple (but already interesting) concentration bound:

LEMMA 2.2. (POOR MAN'S CHERNOFF BOUND) *For $k \geq 1$, $\Pr[X \geq k\sqrt{n}] = 2^{-\Omega(k)}$.*

Proof. For $j \geq 0$, let t_j be the smallest index such that $\sum_{i=1}^{t_j} X_i \geq j \cdot (2\sqrt{n} + 1)$, if such an index exists. For $j \geq 1$, if t_j exists, then $\sum_{i=t_{j-1}+1}^{t_j} X_i \geq 2\sqrt{n}$. So, if we condition on t_{j-1} existing, we can apply Lemma 2.1 to $X_{t_{j-1}+1}, \dots, X_n$ to get $\Pr[t_j \text{ exists} \mid t_{j-1} \text{ exists}] \leq \frac{1}{2}$. By induction on j , this implies $\Pr[t_j \text{ exists}] \leq 2^{-\Omega(j)}$. \square

In addition to the result above, we will need a Chernoff bound for sums of geometric random variables.

LEMMA 2.3. (CHERNOFF BOUND FOR GEOMETRIC R.V.S) *Let Y_1, Y_2, \dots, Y_n be independent real-valued random variables and let $p \in (0, 1)$. If each Y_i satisfies $\Pr[Y_i \geq j] \leq p^j$ for all $j \in \mathbb{N}$, then $Y = \sum_i Y_i$ satisfies $\Pr[Y \geq 2n] \leq (4p)^n$.*

Proof. If $\sum Y_i \geq 2n$ then $\sum [Y_i] \geq n$. Thus there must exist $\vec{a} = (a_1, a_2, \dots, a_n) \in (\mathbb{N} \cup \{0\})^n$ such that $\sum_i a_i = n$ and such that $\max(Y_i, 0) \geq a_i$ for each $i \in [n]$. Let A denote the set of possible vectors \vec{a} . For a given $\vec{a} \in A$,

$$\Pr[\max(Y_i, 0) \geq a_i \text{ for all } i] \leq \prod_i \Pr[\max(Y_i, 0) \geq a_i] \leq \prod_i p^{a_i} = p^{\sum_i a_i} = p^n.$$

By a union bound,

$$\Pr[Y \geq 2n] \leq \sum_{\vec{a} \in A} \Pr[Y_i \geq a_i \text{ for all } i] \leq |A| \cdot p^n.$$

To complete the proof, it suffices to prove $|A| \leq 4^n$. We can encode each $\vec{a} \in A$ as a binary string given by a_1 zeros followed by a one, then a_2 zeros followed by a one, and so on. The string has $\sum_i a_i = n$ zeros and n ones, making for a length of $2n$. We therefore have $|A| \leq 2^{2n} = 4^n$, as desired. \square

Finally, combining the previous lemmas in the right way, we can extract the full bound:

THEOREM 2.1. (CHERNOFF BOUND FOR FAIR COIN FLIPS) *For $k \geq 1$, $\Pr[X \geq k\sqrt{n}] \leq 2^{-\Omega(k^2)}$.*

Proof. Break the coins into k^2 groups of size $n/k^2 \pm 1$ each, and define Y_1, Y_2, \dots, Y_{k^2} so that Y_i is the sum of the X_i s in group i . By Lemma 2.2, we have

$$\Pr \left[Y_i \geq j\sqrt{n/k^2} \right] \leq 2^{-\Omega(j)}.$$

Thus there exists a positive constant c such that $Y'_i := Y_i / (c\sqrt{n/k^2}) = Y_i / (c\sqrt{n}/k)$ satisfies $\Pr[Y'_i \geq j] \leq 8^{-j}$. The Y'_i s are independent geometric random variables, so we can apply Lemma 2.3 (with $p = 1/8$) to get

$$\Pr \left[\sum_{i=1}^{k^2} Y'_i \geq 2k^2 \right] \leq 2^{-\Omega(k^2)}.$$

Plugging in $\sum X_i = \Theta(\sqrt{n}/k) \cdot \sum Y'_i$ proves the theorem. \square

3 Fair Coin Flips: The Same Proof But With Commentary

We will now repeat the proof in the previous section, but this time with ample additional commentary. The goal is to add flavor and intuition to the proof. Along the way, we will also prove a matching lower bound, concluding that $\Pr[X \geq k\sqrt{n}]$ is not just $2^{-\Omega(k^2)}$, but is actually $2^{-\Theta(k^2)}$. To simplify the exposition in this section, we will often ignore rounding errors when discussing division and square roots—alternatively, so that these rounding errors do not exist, you can feel free to imagine that we are focusing only on values of n that are powers of four and k that are powers of 2.

As before, let X_1, X_2, \dots, X_n be independent random coin flips, where $X_i = 1$ represents heads and $X_i = -1$ represents tails. Each X_i independently satisfies $\Pr[X_i = -1] = \Pr[X_i = 1] = 0.5$. Let $X = \sum_i X_i$ count the total number of heads minus the total number of tails. We want to prove the following:

THEOREM 3.1. (Chernoff Bound for Fair Coin Flips) *For $k \in \{1, \dots, \sqrt{n}\}$,*

$$(3.4) \quad \Pr[X \geq k\sqrt{n}] = 2^{-\Theta(k^2)}.$$

We will present the proof of Theorem 3.1 in four bite-sized pieces. The first three pieces can be viewed as warm-up results, each of which has a very simple (almost straightforward) proof. Then, in the final piece, we will show how to combine the warm-up results in order to get the full theorem.

The first warm-up establishes what we call the *Poor Man's Chernoff Bound*. This bound gets the wrong dependence on k , but it will be *incredibly simple to prove*. Moreover (and perhaps surprisingly) the bound will play an important role in the proof of the full theorem.

PROPOSITION 3.1. (Poor Man's Chernoff Bound) *For even $k \leq \sqrt{n}$,*

$$(3.5) \quad \Pr[X \geq k\sqrt{n}] \leq 2^{-k/2}.$$

The second warm-up establishes a very simple Chernoff bound for geometric random variables. This bound might seem like a niche special case, but we will see that it is actually a critical building block for getting tight Chernoff bounds (no matter what parameter regime you care about).

PROPOSITION 3.2. (Sum of Geometric Random Variables) *Let Y_1, Y_2, \dots, Y_n be independent real-valued random variables and let $p \in (0, 1)$. Suppose each Y_i satisfies for all non-negative integers j that*

$$(3.6) \quad \Pr[Y_i \geq j] \leq p^j.$$

Then the sum $Y = \sum_i Y_i$ satisfies

$$\Pr[Y \geq 2n] \leq (4p)^n.$$

The third warm-up result establishes the lower-bound side of Theorem 3.1:

PROPOSITION 3.3. (Fair Coins Lower Bound) *For $k \leq \sqrt{n}$,*

$$(3.7) \quad \Pr[X \geq k\sqrt{n}] \geq 2^{-O(k^2)}.$$

Each warm-up individually has a very simple combinatorial proof. On the other hand, once we have completed the warm-ups, the full proof of Theorem 3.1 will be *almost immediate*. This final part, where we put the pieces together to get the full theorem, is my favorite part of the proof.

3.1 The Poor Man’s Chernoff Bound

Our first warm-up is to prove Proposition 3.1.

PROPOSITION 3.1. (Poor Man’s Chernoff Bound) For even $k \leq \sqrt{n}$,

$$(3.5) \quad \Pr[X \geq k\sqrt{n}] \leq 2^{-k/2}.$$

We will make use of one basic fact:

$$(3.8) \quad \Pr[X \geq 2\sqrt{n}] \leq \frac{1}{4},$$

which follows directly from Chebyshev’s inequality. If you don’t have Chebyshev’s inequality in cache, you can also feel free to take (3.8) as a black-box fact.

Proof. [Proof of Poor Man’s Chernoff Bound] Suppose we flip the n coins one after another, so that X_i gets revealed at time i . Say that we have achieved an *upper deviation* of R at time t if $\sum_{i=1}^t X_i = R$. We will be interested in the *checkpoints* at which we first achieve upper deviations of $2\sqrt{n}$, $4\sqrt{n}$, $6\sqrt{n}$, etc. That is, for $s = 1, 2, \dots$, define the checkpoint t_s to be the earliest point in time at which we have achieved an upper deviation of $2s\sqrt{n}$. See Figure 1.

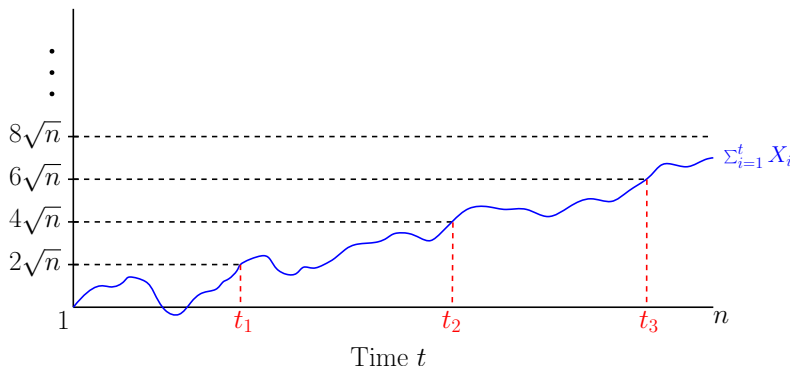


Figure 1: A graph of $\sum_{i=1}^t X_i$ over time t , with labels for the times t_1, t_2, t_3 at which we first achieve upper deviations $2\sqrt{n}$, $4\sqrt{n}$, and $6\sqrt{n}$, respectively. The time t_4 does not exist in this example, because an upper deviation of $8\sqrt{n}$ is never achieved.

Notice that, *a priori*, the checkpoint t_1 may not exist. (We may flip all n coins and never get an upper deviation of $2\sqrt{n}$). Even if t_1 exists, t_2 may not exist. And even if t_2 exists, t_3 may not, etc. Of course, if $X \geq k\sqrt{n}$ then the checkpoint $t_{k/2}$ *must exist* (although the converse is not true). Thus, in order to bound $\Pr[X \geq k\sqrt{n}]$, we can instead bound $\Pr[t_{k/2} \text{ exists}]$.

Let’s begin by proving that $\Pr[t_1 \text{ exists}] \leq 1/2$. Observe that

$$\Pr[X \geq 2\sqrt{n}] = \Pr[t_1 \text{ exists}] \cdot \Pr[X_{t_1+1} + \dots + X_n \geq 0 \mid t_1 \text{ exists}].$$

The probability on the left side is at most $1/4$ by (3.8), and the second probability on the right side is at least $1/2$ by symmetry between heads/tails. Thus $1/4 \geq \Pr[t_1 \text{ exists}] \cdot 1/2$, implying that $\Pr[t_1 \text{ exists}] \leq 1/2$.

Next we argue that $\Pr[t_i \text{ exists} \mid t_1, \dots, t_{i-1} \text{ exist}] \leq 1/2$ for any $i > 1$. Indeed, t_i occurs only if, starting at time $t_{i-1} + 1$, there is some point in time during the remaining $n - t_{i-1} \leq n$ coin flips at which we have again achieved an (additional) upper deviation of $2\sqrt{n}$. However, we already know from our analysis of $\Pr[t_1 \text{ exists}]$ that any sequence of $\leq n$ coin flips has probability at most $1/2$ of ever achieving upper deviation at least $2\sqrt{n}$. Thus $\Pr[t_i \text{ exists} \mid t_1, \dots, t_{i-1} \text{ exist}] \leq 1/2$.

Putting the pieces together,

$$\Pr[X \geq k\sqrt{n}] \leq \Pr[t_{k/2} \text{ exists}] \leq \prod_{i=1}^{k/2} \Pr[t_i \text{ exists} \mid t_1, \dots, t_{i-1} \text{ exist}] \leq \frac{1}{2^{k/2}}.$$

□

It's worth taking a moment to understand the moral of the Poor Man's Chernoff bound. What the bound is really saying is that if we consider thresholds $0, 2\sqrt{n}, 4\sqrt{n}, 6\sqrt{n}, \dots$ for X , the marginal probability of getting from the i -th threshold to the $(i+1)$ -st is decreasing as a function of i . That is, the first upper deviation of $2\sqrt{n}$ is the easiest (occurring with probability roughly $1/2$). The next $2\sqrt{n}$ is the next easiest, and so on. This is an almost trivial fact (since each subsequent deviation has fewer coin flips to make use of than the previous ones), and, as we will see later on, it is also a fact that holds in many settings (not just coin flips). But even this simple fact is enough to get a nontrivial bound.

3.2 A Simple Chernoff bound for Sums of Geometric Random Variables Our next warm-up is to prove Proposition 3.2.

PROPOSITION 3.2. (Sum of Geometric Random Variables) *Let Y_1, Y_2, \dots, Y_n be independent real-valued random variables and let $p \in (0, 1)$. Suppose each Y_i satisfies for all non-negative integers j that*

$$(3.6) \quad \Pr[Y_i \geq j] \leq p^j.$$

Then the sum $Y = \sum_i Y_i$ satisfies

$$\Pr[Y \geq 2n] \leq (4p)^n.$$

Proof. If $Y \geq 2n$, then $Y' = \sum_i \lfloor Y_i \rfloor$ must be at least n . Thus, there exists a tuple of non-negative integers $\langle q_1, q_2, \dots, q_n \rangle$ such that $\sum_i q_i = n$ and such that $\max(Y_i, 0) \geq q_i$ for each i . Call such a tuple a **witness sequence**.

We will complete the proof in two pieces. First, we bound the number of possible witness sequences by 4^n . Next, we bound the probability of a given witness sequence occurring by p^n . Combining these facts, we have by a union bound that the probability of any witness sequence occurring is at most $4^n p^n \leq (4p)^n$.

To bound the number of possible witness sequences, observe that each witness sequence $\langle q_1, q_2, \dots, q_n \rangle$ can be viewed as a way to throw n balls into n bins (i.e., place q_i balls into each bin i). There is a classic trick for bounding the number of ways to do this: encode the witness sequence as a binary string with n zeros and n ones, where the string consists of q_1 ones, followed by a zero, then q_2 ones, followed by a zero, then q_3 ones, followed by a zero, and so on. This creates an injection from witness sequences to binary strings of length $2n$. Since there are trivially at most $2^{2n} = 4^n$ such binary strings, it follows that there are also at most 4^n possible witness sequences.

To bound the probability of a given witness sequence occurring, we can simply apply (3.6). This tells us that each Y_i has probability at most p^{q_i} of satisfying $Y_i \geq q_i$. As the Y_i s are independent, it follows that

$$(3.9) \quad \Pr[Y_i \geq q_i \text{ for all } i] \leq \prod_{i=1}^n \Pr[Y_i \geq q_i] \leq \prod_{i=1}^n p^{q_i} = p^n,$$

where the final equality makes use of the fact that $\sum_{i=1}^n q_i = n$. This completes the proof. \square

Note that, if the Y_i s are guaranteed to be integers, then the preceding argument gives us a slightly stronger bound (since we can use Y in place of Y').

COROLLARY 3.1. *If the Y_i s are guaranteed to be integers, then*

$$\Pr[Y \geq \mathbb{E}[Y] + n] \leq (4p)^n.$$

3.3 The Lower Bound Our third warm-up is to prove the lower-bound side of our Chernoff bound. Although we don't typically prove the lower-bound side when we teach Chernoff bounds, we will see that its proof is *remarkably* simple.

PROPOSITION 3.3. (Fair Coins Lower Bound) *For $k \leq \sqrt{n}$,*

$$(3.7) \quad \Pr[X \geq k\sqrt{n}] \geq 2^{-O(k^2)}.$$

To prove (3.7), we will make use of another basic fact:

$$(3.10) \quad \Pr[X \geq \sqrt{n}/4] \geq 1/4.$$

To streamline our exposition, we will take (3.7) for granted. For completeness, however, we also include a simple combinatorial proof in Appendix A.

Proof. [Proof of Proposition 3.3] Partition the coins into k^2 groups each of size $S = n/k^2$. Define E_i to be the event that group i achieves sum of at least $\sqrt{S}/4 = \sqrt{n/k^2}/4 = \sqrt{n}/(4k)$. Notice that, if *all* of events E_1, E_2, \dots, E_{k^2} were to occur, then the total sum would be at least $k^2 \cdot \sqrt{n}/(4k) \geq k\sqrt{n}/4$. See Figure 2.

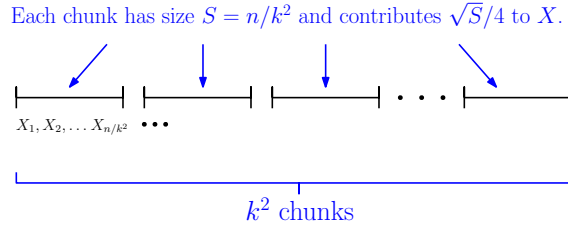


Figure 2: The lower-bound construction partitions the coins into k^2 groups, and considers the event that every group contributes $\Omega(\sqrt{S})$ to X , where S is the size of each group. This would imply that $X \geq k^2 \cdot \Omega(\sqrt{S}) = \Omega(k^2 \cdot \sqrt{n/k^2}) = \Omega(k\sqrt{n})$.

Applying (3.10) to each group i (with $n = |S|$), we see that each E_i occurs with probability at least $1/4$. The probability that all of E_1, E_2, \dots, E_{k^2} occur is therefore at least $1/4^{k^2}$. Thus we have that

$$(3.11) \quad \Pr[X \geq k\sqrt{n}/4] \geq 1/4^{k^2}.$$

This is not quite what we set out to prove, since we want $\Pr[X \geq k\sqrt{n}]$. Notice, however, that by a simple change of variables, (3.11) implies that $\Pr[X \geq k\sqrt{n}] \geq (1/4)^{16k^2}$ for all $k \leq \sqrt{n}/4$. And the case of $k \geq \sqrt{n}/4$ follows from the simple fact that, for $k = \Theta(\sqrt{n})$, we have $\Pr[X \geq k\sqrt{n}] \geq \Pr[X = n] = 2^{-n} = 2^{-O(k^2)}$. \square

It is important to understand why we broke the coins into k^2 groups, rather than some other number: k^2 is the magic number such that, if each group misbehaves just a *little* (i.e., incurs a sum of at least $0.1\sqrt{S}$, which occurs with probability at least 0.1), then the cumulative effect is a sum of $\Omega(k\sqrt{n})$.

3.4 The Upper Bound

We are now prepared to prove the full Chernoff bound, restated below.

THEOREM 3.2. (Chernoff Bound for Fair Coin Flips) *For $k \in \{1, \dots, \sqrt{n}\}$,*

$$(3.4) \quad \Pr[X \geq k\sqrt{n}] = 2^{-\Theta(k^2)}.$$

Proof. As we have already proven the lower bound (Proposition 3.3), we can focus here on the upper bound. We will show that

$$(3.12) \quad \Pr[X \geq 16k\sqrt{n}] \leq \frac{1}{4^{k^2}}.$$

Our proof will build on each of the three warm-ups from earlier. We will use essentially the same group structure as in the lower bound, and we will analyze the groups by *directly applying* Propositions 3.1 and 3.2.

Break the coins into k^2 groups, each of size $S = n/k^2$. Define C_1, C_2, \dots, C_{k^2} to be the sums of the coin flips in each group. One way to think about the event $X \geq 16k\sqrt{n}$ is that the C_i s are, on average, at least $16k\sqrt{n}/k^2 = 16\sqrt{S}$. We know that for each group, having a sum of $16\sqrt{S}$ isn't very likely—in fact, by the Poor Man's Chernoff Bound (Proposition 3.1), we know that each C_i satisfies the bound

$$\Pr[C_i \geq 8t\sqrt{S}] \leq 2^{-4t}$$

for any positive integer t . If we define $\bar{C}_i = C_i/(8\sqrt{S})$, then the Poor Man's bound translates to

$$\Pr[\bar{C}_i \geq t] \leq 1/16^t.$$

In other words, \bar{C}_i is bounded above by a geometric random variable.

We are interested in the event that

$$\sum_{i=1}^{k^2} C_i \geq 16k\sqrt{n}.$$

As noted above, this is equivalent to the event that, on average, each C_i is at least $16\sqrt{S}$. Rewriting this in terms of the \bar{C}_i s, the event that we care about is

$$\sum_{i=1}^{k^2} \bar{C}_i \geq 2k^2.$$

Since the \bar{C}_i s are independent geometric random variables, we can apply Proposition 3.2 to bound the probability of the above event by

$$(4/16)^{k^2} = 4^{-k^2},$$

which completes the proof. \square

4 Biased Coin Flips: The Large-Deviations Case

Next we extend our Chernoff bound to handle biased coin flips. Let $p \leq 1/2$ be a probability. Suppose that each of X_1, X_2, \dots, X_n is 0 with probability $1 - p$ and 1 with probability p . As before, assume that the X_i s are independent, and set $X = \sum_i X_i$. Notice that we have swapped from each X_i being in $\{-1, 1\}$ to each X_i being in $\{0, 1\}$. This perspective, it turns out, will significantly simplify the exposition when we present the analysis for the large-deviation regime.

Let $\mu = \mathbb{E}[X] = pn$. As discussed in the introduction, the Chernoff bound for X splits into two parameter regimes. The *small-deviation* regime is governed by a bound that looks very similar to what we had for fair coin flips: for $k \in \{1, 2, \dots, \sqrt{\mu}\}$,

$$(4.13) \quad \Pr[X \geq \mu + k\sqrt{\mu}] = 2^{-\Theta(k^2)}.$$

The *large-deviation* regime is governed by a bound that looks a little different: for any $r \geq 1$ satisfying $r\mu \leq n$, we have

$$(4.14) \quad \Pr[X \geq r\mu] = 1/\Theta(r)^{\Theta(r\mu)}.$$

The small-deviation case follows from almost exactly the same arguments as in the previous section, so we will skip its proof for now. (But, for completeness, it is worth noting that Theorems 5.1 and 5.2 in Section 5 directly imply (4.13).) Instead, this section will focus on the large-deviation regime. What's neat is that the proof of (4.14) will follow almost exactly the same structure as the proof that we have already seen.

PROPOSITION 4.1. *Let $0 \leq p \leq 1/2$. Let X_1, \dots, X_n be i.i.d. 0-1 random variables satisfying $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$. Let $X = \sum_i X_i$ and let $\mu = \mathbb{E}[X] = pn$. For any $r \geq 1$ satisfying $r\mu \leq n$, we have*

$$\Pr[X \geq r\mu] = 1/\Theta(r)^{\Theta(r\mu)}.$$

To simplify our discussion, we will assume in our proof of Proposition 4.1 that n is divisible by $r\mu$. This is just to avoid some minor handling of rounding errors, and with a bit of casework one can actually show that this simplification is without loss of generality.

Proof. We begin with the lower bound. Break the coins into $r\mu$ groups. The coins in each group have cumulative expectation $p \cdot \frac{n}{r\mu} = p \cdot \frac{n}{rpn} = \frac{1}{r}$. With a bit of work, one can obtain the following basic fact: the probability of at least one coin in the group evaluating to 1 is at least $\Omega(1/r)$.¹ It follows that, with probability at least $\Omega(1/r)^{r\mu}$, every group will contribute at least 1 to X , making for a total of at least $r\mu$. Thus

$$\Pr[X \geq r\mu] \geq \Omega(1/r)^{r\mu}.$$

This establishes the lower-bound direction.

To derive the upper bound, we need to first derive something that closely resembles the Poor Man's Chernoff Bound for the X_i s in a given group. Let X_a, \dots, X_b be the X_i s that comprise some group, and let $C = \sum_{i=a}^b X_i$. Since $\mathbb{E}[C] = 1/r$, we know from Markov's inequality that $\Pr[C \geq 1] \leq 1/r$. That is, if we flip the coins X_a, \dots, X_b one after another, the probability that we ever get a 1 is at most $1/r$. Similarly, once we get that 1, the probability that we ever get *another* 1 in the same group is again at most $1/r$. Continuing like this, we can conclude that

$$(4.15) \quad \Pr[C \geq k] \leq r^{-k}.$$

Now we can complete the proof using Proposition 3.2 (or, since the X_i s are integers, we can actually use Corollary 3.1). Define $C_1, C_2, \dots, C_{r\mu}$ so that C_i is the sum of the X_i s in the i -th group. Equation (4.15) tells us that each C_i is bounded above by a geometric random variable. It follows by Corollary 3.1 that

$$\Pr \left[\sum_{i=1}^{r\mu} C_i \geq r\mu \right] \leq (4/r)^{r\mu}.$$

Since $\sum_i X_i = \sum_i C_i$, this completes the proof of the upper bound.

So, by following almost exactly the same template as before, we once again arrive at nearly matching upper and lower bounds. \square

5 Generalizing to Bennett's Inequality

Part of what makes the proof approach in this paper useful is that the basic approach naturally extends to many other settings. To showcase, this, we will prove in this section an adaptive version of Bennett's inequality [2]. More generally, the proof also extends to give Azuma's inequality [1], McDiarmid's inequality [22], and various Bernstein-type inequalities [3, 4] (and, indeed, in their most basic formulations, all of these inequalities are corollaries of Theorem 5.1, hence our focus on Bennett's inequality).

THEOREM 5.1. (ADAPTIVE VERSION OF BENNETT'S INEQUALITY) *Let $n \in \mathbb{N}$ and $v \in \mathbb{R}^+$. Suppose that Alice selects $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$ where each \mathcal{D}_i is a probability distribution over $[-\infty, 1]$ with mean 0 and with some variance v_i . Alice selects $\mathcal{D}_1, \mathcal{D}_2, \dots$ one at a time, and once a given \mathcal{D}_i is selected, a random variable X_i is drawn from the distribution \mathcal{D}_i . Alice gets to select the \mathcal{D}_i s (and thus also the v_i s) adaptively, basing \mathcal{D}_i on the outcomes of X_1, \dots, X_{i-1} . The only constraint on Alice is that $\sum_{i=1}^n v_i \leq v$.*

Define $X = \sum_{i=1}^n X_i$. Then, for $k \in [1, \sqrt{v}]$, we have that

$$\text{(the small-deviation case)} \quad \Pr[X \geq k\sqrt{v}] \leq 2^{-\Omega(k^2)}.$$

And for $r \geq 1$, we have

$$\text{(the large-deviation case)} \quad \Pr[X \geq rv] \leq O(1/r)^{\Omega(rv)}.$$

In the small-deviation case, and with a few extra constraints on Alice (namely that $X_i \in [-1, 1]$ and that $\sum v_i = v$), we can also get a matching lower bound. As far as I know, this lower bound has not appeared in past work, and does not follow from standard techniques.

¹Indeed, the probability that exactly one coin evaluates to 1 is $\binom{n}{1} p(1-p)^{n-1}$. Since $n/(r\mu) = p^{-1}r^{-1}$, this probability is at least $r^{-1}(1-p)^{p^{-1}-1} \geq r^{-1}/e$.

THEOREM 5.2. (LOWER BOUND FOR BENNETT'S INEQUALITY) Let $n, v \in \mathbb{N}$. Suppose that Alice selects $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$ where each \mathcal{D}_i is a probability distribution over $[-1, 1]$ with mean 0 and with some variance v_i . Alice selects $\mathcal{D}_1, \mathcal{D}_2, \dots$ one at a time, and once a given \mathcal{D}_i is selected, a random variable X_i is drawn from the distribution \mathcal{D}_i . Alice gets to select the \mathcal{D}_i s (and thus also the v_i s) adaptively, basing \mathcal{D}_i on the outcomes of X_1, \dots, X_{i-1} . The only constraint on Alice is that $\sum_{i=1}^n v_i = v$.

Define $X = \sum_{i=1}^n X_i$. Then, for $k \in [1, \sqrt{v}]$, we have that

$$\Pr[X \geq \Omega(k\sqrt{v})] \geq 2^{-O(k^2)}.$$

It is worth noting that, as an immediate corollary of Theorem 5.1, we also get a general-purpose Chernoff bound for non-iid real-valued coin flips (sometimes known as Hoeffding's bound).

COROLLARY 5.1. (CHERNOFF BOUND FOR NON-IDENTICAL REAL-VALUED COIN FLIPS [15]) Let $X_1, \dots, X_n \in [0, 1]$ be independent random variables with means p_1, \dots, p_n . Let $\mu = \sum_i p_i$ and let $X = \sum X_i$. Then, for any integer $k \leq \sqrt{\mu}$,

$$(5.16) \quad \Pr[X \geq \mu + k\sqrt{\mu}] \leq 2^{-\Omega(k^2)}.$$

And for any $r \geq 1$,

$$\Pr[X \geq \mu + r\mu] \leq O(1/r)^{\Omega(r\mu)}.$$

Proof. Each X_i has variance $\mathbb{E}[X_i^2] - p_i^2 \leq \mathbb{E}[X_i] \leq p_i$. So the result follows from Theorem 5.1. \square

The rest of this section is structured as follows. Section 5.1 presents some basic preliminaries, culminating in a variation of the Poor Man's Chernoff Bound that can be used in the proof Theorem 5.2. Sections 5.2 and 5.3 then prove the small and large deviation cases, respectively, for Theorem 5.1. Finally, Section 5.4 proves Theorem 5.2.

5.1 Preliminaries We begin by proving some preliminary lemmas that will be useful for both the upper and lower bounds. Our first lemma bounds the variance of X by v .

LEMMA 5.1. We have $\mathbb{E}[X^2] = \mathbb{E}[\sum_i v_i] \leq v$.

Proof. Define $Y_i = \sum_{j=1}^i X_j$. It suffices to show that for each $i \in [n]$, we have $\mathbb{E}[Y_i^2 - Y_{i-1}^2] = \mathbb{E}[v_i]$. By linearity of expectation,

$$\mathbb{E}[Y_i^2] = \mathbb{E}[(X_i + Y_{i-1})^2] = \mathbb{E}[X_i^2] + \mathbb{E}[Y_{i-1}^2] + 2\mathbb{E}[X_i Y_{i-1}].$$

No matter the outcome of Y_{i-1} , we have that $\mathbb{E}[X_i] = 0$, so $\mathbb{E}[X_i Y_{i-1}] = 0$. Thus $\mathbb{E}[Y_i^2] = \mathbb{E}[X_i^2] + \mathbb{E}[Y_{i-1}^2] = \mathbb{E}[v_i] + \mathbb{E}[Y_{i-1}^2]$, as desired. \square

Let $Y_{\max} = \max_j \sum_{i=1}^j X_i$ be the largest sum achieved by any prefix of the X_i s. Our next lemma uses Chebyshev's inequality to bound Y_{\max} .

LEMMA 5.2. For any $\ell \geq 1$, $\Pr[Y_{\max}^2 \geq \ell v] \leq \frac{1}{\ell}$.

Proof. By Lemma 5.1, we know that the variance of X is at most v . By Chebyshev's inequality, it follows that $\Pr[X^2 \geq \ell v] \leq \frac{1}{\ell}$.

Notice, however, that if $Y_{\max}^2 \geq \ell v$, then Alice can also force $X^2 \geq \ell v$: as soon as $(\sum_{i=1}^j X_i)^2 \geq \ell v$ for some j , she simply sets X_{j+1}, \dots, X_n to be deterministically 0. Thus any tail bound on X implies the same tail bound on Y_{\max} , and the lemma is proven. \square

Our next lemma establishes the key technical ingredient that we will need in order to obtain a Poor Man's Chernoff Bound. The lemma transforms bounds on $\Pr[Y_{\max} \geq \alpha]$, for a given α , into bounds on $\Pr[Y_{\max} \geq \alpha + \beta + 1 \mid Y_{\max} \geq \beta]$ for a given α, β .

LEMMA 5.3. (LAW OF DIMINISHING GROWTH) *Let q be a real number such that $\Pr[Y_{\max} \geq \alpha] \leq q$ holds independently of Alice's strategy. For any $\beta > 0$, if Alice follows a strategy \mathcal{A} such that $\Pr[Y_{\max} \geq \beta] > 0$, then $\Pr[Y_{\max} \geq \beta + \alpha + 1 \mid Y_{\max} \geq \beta] \leq q$.*

Proof. Suppose Alice follows strategy \mathcal{A} and that $Y_{\max} \geq \beta$. Let j be the smallest j such that $\beta \leq \sum_{i=1}^j X_i \leq \beta + 1$. Define $Y'_{\max} = \max_{k > j} \sum_{i=j+1}^k X_i$ to be the maximum sum achieved by any prefix of $X_{j+1}, X_{j+2}, \dots, X_n$.

If $Y_{\max} \geq \beta + \alpha + 1$, then we must have $Y'_{\max} \geq \alpha$. On the other hand, by the definition of q , we have $\Pr[Y'_{\max} \geq \alpha] \leq q$.² Thus $\Pr[Y_{\max} \geq \beta + \alpha + 1 \mid Y_{\max} \geq \beta] \leq q$. \square

Building on Lemma 5.3, we can immediately obtain a Poor Man's Chernoff Bound:

LEMMA 5.4. (POOR MAN'S CHERNOFF BOUND) *If $v \geq 1$, then for any $k \in \mathbb{N}$,*

$$(5.17) \quad \Pr[X \geq 4k\sqrt{v}] \leq 1/4^k.$$

And if $v \leq 1$, then for any $k \in \mathbb{N}$,

$$(5.18) \quad \Pr[X \geq k] \leq v^{k/2}.$$

Proof. If $v \geq 1$, then Lemma 5.2 bounds $\Pr[Y_{\max}^2 \geq 4v] \leq 1/4$ and thus $\Pr[Y_{\max} \geq 2\sqrt{v}] \leq 1/4$. By Lemma 5.3, it follows that for any $j \in \mathbb{N}$, we have $\Pr[Y_{\max} \geq (2\sqrt{v} + 1)j] \leq \frac{1}{4^j}$, and thus that $\Pr[Y_{\max} \geq 4\sqrt{v}j] \leq \frac{1}{4^j}$.

If $v \leq 1$, then Lemma 5.2 bounds $\Pr[Y_{\max}^2 \geq 1] \leq v$ and thus $\Pr[Y_{\max} \geq 1] \leq v$. By Lemma 5.3, it follows that for any integer $j \geq 0$, $\Pr[Y_{\max} \geq 1 + 2j] \leq v^{1+j}$, which implies (5.18). \square

Of course, the bounds achieved in the previous lemma are much weaker than those stated by Theorem 5.1. On the other hand, we achieved them with almost no work: we simply combined a trivial application of Chebyshev's inequality (Lemma 5.2) with a natural observation about sums of random variables (Lemma 5.3).

Finally, we will also need an extension of Proposition 3.2, our Chernoff bound for Geometric Random Variables:

PROPOSITION 5.1. *Let Y_1, Y_2, \dots, Y_n be real-valued random variables and let $p \in (0, 1)$. Suppose that, for each $i \in [n]$ and $j \in \mathbb{N}$, if we condition on any outcomes for Y_1, Y_2, \dots, Y_{i-1} , then Y_i satisfies*

$$\Pr[Y_i \geq j \mid Y_1, \dots, Y_{i-1}] \leq p^j.$$

Then the sum $Y = \sum_i Y_i$ satisfies

$$\Pr[Y \geq 2n] \leq (4p)^n.$$

Proof. The proof is the same as for Proposition 3.2, except that the reasoning which we use to derive (3.9) now becomes:

$$\Pr[Y_i \geq q_i \text{ for all } i] \leq \prod_{i=1}^n \Pr[Y_i \geq q_i \mid Y_1 \geq q_1, \dots, Y_{i-1} \geq q_{i-1}] \leq \prod_{i=1}^n p^{q_i} = p^n.$$

\square

With these preliminaries in place, we can now proceed to prove much stronger bounds using the same approach as in previous sections.

²Here, we are implicitly using the following observation: any strategy that Alice can use to make Y'_{\max} large could also be used to make Y_{\max} large, as Alice can set $X_1, X_2, \dots, X_j := 0$ in order to force $Y'_{\max} = Y_{\max}$.

5.2 Upper Bound for the Small-Deviation Regime In this section, we establish an upper bound for the small-deviation regime: we prove that for any $k \leq \sqrt{v}$, we have

$$(5.19) \quad \Pr[X \geq 33k\sqrt{v}] \leq \frac{1}{4k^2}.$$

Call an X_i *oversized* if $v_i \geq v/k^2$. Since $\sum_i v_i \leq v$, there can be at most k^2 oversized X_i s. Since each X_i satisfies $X_i \leq 1$, the total contribution of oversized X_i s to X is at most $k^2 \leq k\sqrt{v}$. In the rest of the proof, we will assume without loss of generality that there are no oversized X_i s and that our task is to bound $\Pr[X \geq 32k\sqrt{v}]$.

Partition the random variables X_1, X_2, \dots, X_n into at most k^2 *groups* such that the X_i s in each group have variances that sum to at most $2v/k^2$. We define the partition greedily: we end group j and begin group $j+1$ once the X_i s in group j have sum of variances at least v/k^2 . Note that, since the v_i s are random variables, even the outcome of which X_i s are in each group are random variables.

Define C_1, C_2, \dots, C_{k^2} so that C_i is the sum of the X_j s in the i -th group (or 0 if no such group exists). Each C_i has expected value 0, variance at most $2v/k^2$ (by Lemma 5.1), and standard deviation at most $\sqrt{2v/k^2} \leq 2\sqrt{v}/k$.

Throughout the rest of the proof, define $\sigma = 2\sqrt{v}/k$ to be an upper bound on the standard deviation of each C_i . The statement $X \geq 32k\sqrt{v}$ is equivalent to the statement $X \geq 16k^2\sigma$. Thus our goal is to bound

$$(5.20) \quad \Pr[X \geq 16k^2\sigma] \leq \frac{1}{4k^2}.$$

One should think of this as the probability that the average C_i is at least 16 standard deviations large.

Condition on any outcomes for C_1, \dots, C_{i-1} . Then, applying the Poor Man's Chernoff Bound (Lemma 5.4) to C_i , we have that

$$\Pr[C_i \geq 4t\sigma \mid C_1, \dots, C_{i-1}] \leq 1/4^t.$$

Rewriting this in terms of $\bar{C}_i := C_i/(8\sigma)$ gives

$$\Pr[\bar{C}_i \geq t \mid \bar{C}_1, \dots, \bar{C}_{i-1}] \leq 1/16^t.$$

In other words, $\bar{C}_i \mid \bar{C}_1, \dots, \bar{C}_{i-1}$ is bounded above by a geometric random variable (with mean roughly $1/16$), and this holds no matter the outcomes of C_1, \dots, C_{i-1} .

It follows by Proposition 5.1 that

$$\Pr \left[\sum_{i=1}^{k^2} \bar{C}_i \geq 2k^2 \right] \leq 4^{-k^2}.$$

Rewriting this in terms of the C_i s gives (5.20), as desired.

5.3 Upper Bound for the Large-Deviation Regime In this section, we establish an upper bound for the large-deviation regime: we prove that for any $r \geq 1$ and for any $v \geq 0$ such that rv is a positive integer,

$$(5.21) \quad \Pr[X \geq 3rv] \leq \frac{1}{(32/r)^{rv/2}}.$$

This implies the large-deviation case in Theorem 5.1.

Call an X_i *oversized* if $v_i \geq 1/r$. Since $\sum_i v_i \leq v$, there can be at most rv oversized X_i s. Since each X_i satisfies $X_i \leq 1$, the total contribution of oversized X_i s to X is at most rv . In the rest of the proof, we will assume without loss of generality that there are no oversized X_i s and that our task is to bound $\Pr[X \geq 2rv]$.

Partition the random variables X_1, X_2, \dots, X_n into at most rv *groups* such that the X_i s in each group have variances that sum to at most $\frac{2}{r}$. (Note that, by Lemma 5.1, applied just to the X_i s in the group, this implies that the sum of the X_i s in the group has variance at most $\frac{2}{r}$.) To construct the groups, we define the partition greedily, meaning that we end group j and begin group $j+1$ once group j has sum of variances at least $\frac{1}{r}$. The final X_i in the group has variance $v_i \leq 1/r$ by assumption, so the sum of the variances in each group is at most $2/r$.

Define C_1, C_2, \dots, C_{rv} so that C_i is the sum of the X_j s in the i -th group (or 0 if no such group exists), conditioned on the outcomes of the previous groups C_1, C_2, \dots, C_{i-1} . Each C_i has expected value 0 and variance at most $2/r$ (by Lemma 5.1).

Condition on any outcomes for C_1, \dots, C_{i-1} . By our Poor Man's Chernoff Bound (Lemma 5.4) to C_i , we have that

$$\Pr[C_i \geq t \mid C_1, \dots, C_{i-1}] \leq (2/r)^{t/2}.$$

In other words, $C_i \mid C_1, \dots, C_{i-1}$ is bounded above by a geometric random variable with mean $\Theta(\sqrt{1/r})$, and this holds no matter the outcomes of C_1, \dots, C_{i-1} .

It follows by Proposition 5.1 that

$$\Pr \left[\sum_{i=1}^{rv} C_i \geq 2rv \right] \leq (4\sqrt{2/r})^{rv} \leq (32/r)^{rv/2}.$$

This completes the proof of (5.21).

5.4 Lower Bound for the Small-Deviation Regime Finally, we prove Theorem 5.2. That is, we wish to show that, if Alice is required to satisfy $\sum_i v_i = v$ (rather than just $\sum_i v_i \leq v$), and if Alice is required to guarantee that each $X_i \in [-1, 1]$ (rather than $(-\infty, 1]$), then for any $k \leq \sqrt{v}$, we have

$$(5.22) \quad \Pr[X \geq \Omega(k\sqrt{v})] \geq \frac{1}{2^{O(k^2)}}.$$

Partition the random variables X_1, X_2, \dots, X_n into $\Theta(k^2)$ **groups** such that the variances of the X_i 's in each group sum to between v/k^2 and $2v/k^2$. We define the partition greedily, meaning that once a given group j reaches v/k^2 , and if the total amount of remaining variance is at least v/k^2 , then we begin a new group $j+1$.

Define $C_1, C_2, \dots, C_{\Theta(k^2)}$ so that C_i is the sum of the X_j s in the i -th group. Each C_i has expected value 0, variance at least v/k^2 (by Lemma 5.1), and standard deviation at least \sqrt{v}/k .

We will prove the following lemma:

LEMMA 5.5. *No matter the outcomes of C_1, \dots, C_{i-1} , we have*

$$\Pr[C_i \geq \Omega(\sqrt{v}/k)] \geq \Omega(1).$$

It follows that,

$$\Pr[\text{every } C_i \text{ satisfies } C_i \geq \Omega(\sqrt{v}/k)] \geq \Omega(1)^{O(k^2)} = \frac{1}{2^{O(k^2)}}.$$

On the other hand, if every C_i satisfies $C_i \geq \Omega(\sqrt{v}/k)$, then $X = \sum_{i=1}^{\Theta(k^2)} C_i \geq \Omega(k\sqrt{v})$. Thus, if we can establish Lemma 5.5, then we will have also proven (5.22).

Before diving into the proof of Lemma 5.5, I should make a small apology. The following proof is somewhat more messy than I would like it to be. With that said, the result that we are proving is ancillary—it's only needed for the lower bound. In fact, the lower bound in this subsection is so rarely discussed that I have not been able to find *any* examples of it being discussed or even mentioned in the literature.

Proof. [Proof of Lemma 5.5] We make use of two facts. The first is that C_i has variance at least v/k^2 , so

$$(5.23) \quad \mathbb{E}[C_i^2] \geq v/k^2.$$

The second is that $|C_i|$ is unlikely to be large. For this, Lemma 5.4 suffices, telling us that

$$\Pr[|C_i| \geq 4j\sqrt{\text{Var}(C_i)}] \leq 1/4^j,$$

which, since $\text{Var}(C_i) \leq 2v/k^2 \leq (2\sqrt{v}/k)^2$, means that

$$(5.24) \quad \Pr[|C_i| \geq 8j\sqrt{v}/k] \leq 1/4^j \leq 1/2^j.$$

For notational convenience, define $Q = \frac{C_i}{\sqrt{v}/k}$. Our goal is thus to establish that $Q \geq \Omega(1)$ with probability $\Omega(1)$. Equations (5.23) and (5.24) translate to

$$(5.25) \quad \mathbb{E}[Q^2] \geq 1.$$

and

$$(5.26) \quad \Pr[|Q| \geq 8j] \leq 1/2^j.$$

It might seem strange that (5.26) would be useful in this proof, since (5.26) is an *upper bound* on $|Q|$ and we want a *lower bound* on Q . Importantly, however, (5.26) forces $\mathbb{E}[|Q|]$ and $\mathbb{E}[Q^2]$ to be almost completely determined by cases where $|Q|$ is small. Indeed, two immediate consequences of (5.26) are that, if c is a sufficiently large positive constant, then

$$(5.27) \quad \mathbb{E}[|Q| \cdot \mathbb{I}_{|Q| \geq c}] \leq 1/c$$

and

$$(5.28) \quad \mathbb{E}[Q^2 \cdot \mathbb{I}_{|Q| \geq c}] \leq 1/c.$$

Take c to be a sufficiently large positive constant and suppose for contradiction that $\Pr[|Q| \geq 1/c] \leq 1/c^3$. Then by (5.28),

$$\begin{aligned} \mathbb{E}[Q^2] &\leq \Pr[|Q| \leq 1/c] \cdot 1/c^2 + \Pr[1/c \leq |Q| \leq c] \cdot c^2 + \mathbb{E}[Q^2 \cdot \mathbb{I}_{|Q| \geq c}] \\ &\leq 1/c^2 + 1/c + 1/c < 1, \end{aligned}$$

which contradicts (5.25). Thus $\Pr[|Q| \geq 1/c] \geq \Omega(1)$.

This doesn't complete the proof, because we are interested in $\Pr[Q \geq \Omega(1)]$, not $\Pr[|Q| \geq \Omega(1)]$. However, the fact that $\Pr[|Q| \geq 1/c] \geq \Omega(1)$ does establish that $\mathbb{E}[|Q|] \geq \Omega(1)$. Since $\mathbb{E}[Q] = 0$, it follows that if we define $Q' = Q \cdot \mathbb{I}_{Q \geq 0}$ to be the positive component of Q , then

$$(5.29) \quad \mathbb{E}[Q'] \geq \Omega(1).$$

Now take c' to be a sufficiently large positive constant and suppose for contradiction that $\Pr[Q' \geq 1/c'] \leq 1/c'^2$. Then by (5.27),

$$\begin{aligned} \mathbb{E}[Q'] &\leq \Pr[Q' \leq 1/c'] \cdot 1/c' + \Pr[1/c' \leq Q \leq c'] \cdot c' + \mathbb{E}[|Q| \cdot \mathbb{I}_{Q \geq c'}] \\ &\leq 1/c' + 1/c' + 1/c', \end{aligned}$$

which, if c' is sufficiently large, contradicts (5.29). Thus there exists some positive constant c' such that $\Pr[Q' \geq 1/c'] \geq 1/c'^2 = \Omega(1)$. This establishes that

$$\Pr[Q' \geq \Omega(1)] = \Omega(1),$$

which implies that $\Pr[Q \geq \Omega(1)] = \Omega(1)$ and thus that $\Pr[C_i \geq \Omega(\sqrt{v}/k)] \geq \Omega(1)$. \square

6 Acknowledgments

The author would like to thank Rose Silver for her extensive feedback and comments on early versions of this paper. The author would also like to thank an anonymous reviewer for catching several significant typos and bugs.

This work was partially supported by a Harvard Rabin Postdoctoral Fellowship and by a Harvard FODSI fellowship under NSF grant DMS-2023528.

Parts of this research were completed while William was a PhD student at MIT, where he was funded by a Fannie and John Hertz Fellowship and an NSF GRFP Fellowship. William Kuszmaul was also partially sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

- [1] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.
- [2] George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- [3] Sergei Bernstein. On a modification of Chebyshev’s inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math*, 1(4):38–49, 1924.
- [4] Sergei N Bernstein. On certain modifications of Chebyshev’s inequality. *Doklady Akademii Nauk SSSR*, 17(6):275–277, 1937.
- [5] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [6] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.
- [7] Herman Chernoff. A career in statistics. *Past, Present, and Future of Statistical Science*, 29, 2014.
- [8] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.
- [9] Vasek Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979.
- [10] Victor H de la Pena, Michael J Klass, and Tze Leung Lai. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Annals of Probability*, pages 1902–1933, 2004.
- [11] Kacha Dzhaparidze and JH Van Zanten. On Bernstein-type inequalities for martingales. *Stochastic Processes and Their Applications*, 93(1):109–117, 2001.
- [12] Xiequan Fan, Ion Grama, Quansheng Liu, et al. Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, 20, 2015.
- [13] David A Freedman. On tail probabilities for martingales. *Annals of Probability*, pages 100–118, 1975.
- [14] Erich Haeusler. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(4):523–534, 1984.
- [15] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.
- [16] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 617–631. Springer, 2010.
- [17] Rasul A Khan. Lp-version of the dubins–savage inequality and some exponential inequalities. *Journal of Theoretical Probability*, 22(2):348, 2009.
- [18] William Kuzmaul. Chernoff bounds, part 2: Mechanizing the process. <https://mathandmaking.wordpress.com/2016/11/16/chernoff-bounds-part-2-mechanizing-the-process/>, 2016. Math and Making Blog Post.
- [19] Emmanuel Lesigne and Dalibor Volný. Large deviations for martingales. *Stochastic Processes and Their Applications*, 96(1):143–159, 2001.
- [20] Robert Liptser and Vladimir Spokoiny. Deviation probability bound for martingales with applications to statistical estimation. *Statistics & probability letters*, 46(4):347–357, 2000.
- [21] Quansheng Liu and Frédérique Watbled. Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment. *Stochastic processes and their applications*, 119(10):3101–3132, 2009.
- [22] Colin McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141(1):148–188, 1989.
- [23] Pat Morin, Wolfgang Mulzer, and Tommy Reddad. Encoding arguments. *ACM Computing Surveys (CSUR)*, 50(3):1–36, 2017.
- [24] Wolfgang Mulzer. Five proofs of chernoff’s bound with applications. *arXiv preprint arXiv:1801.03365*, 2018.
- [25] Iosif Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *The Annals of Probability*, pages 1679–1706, 1994.
- [26] Emmanuel Rio et al. Extensions of the Hoeffding–Azuma inequalities. *Electronic Communications in Probability*, 18, 2013.
- [27] Emmanuel Rio et al. On McDiarmid’s concentration inequality. *Electronic Communications in Probability*, 18, 2013.
- [28] Thomas Steinke and Jonathan Ullman. Subgaussian tail bounds via stability arguments. *arXiv preprint arXiv:1701.03493*, 2017.
- [29] Sara A van de Geer. On Hoeffding’s inequality for dependent random variables. In *Empirical Process Techniques for Dependent Data*, pages 161–169. Springer, 2002.
- [30] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

A Proof of (3.7)

In this appendix, we will give a simple combinatorial proof of the following basic lemma about coin flips:

PROPOSITION A.1. *Let X_1, X_2, \dots, X_n be fair and independent ± 1 coin flips. Then, $\Pr[X \geq \sqrt{n}/4] \geq 1/4$.*

Define $Z_i = \sum_{j=1}^i X_j$. We can think of the Z_i s as following a random walk, starting at 0, and progressing ± 1 with equal probability on each step. As a thought experiment, let us continue this random walk in perpetuity, so we extend Z_1, Z_2, \dots, Z_n to an infinite sequence Z_1, Z_2, \dots .

For each $r \geq 1$, let t_r be the first time that the random walk reaches $\pm r$, that is, the smallest $t \geq 1$ such that $|Z_t| = r$. It is a simple exercise to show that t_r exists with probability 1.³

The main step in proving Proposition A.1 is to solve for $\mathbb{E}[t_r]$.

LEMMA A.1. *For every power of two $r = 2^i$, we have that*

$$\mathbb{E}[t_r] = r^2.$$

Proof. We can prove this by induction. Since $t_1 = 1$, it suffices to show that for all $r = 2^i > 1$,

$$(A.1) \quad \mathbb{E}[t_r] = 4\mathbb{E}[t_{r/2}].$$

We can prove (A.1) with a simple thought experiment. Suppose we wish to solve for $\mathbb{E}[t_r]$. The expected time to get from 0 to $\pm r/2$ is just $\mathbb{E}[t_{r/2}]$. Say, without loss of generality, that we get to $r/2$, rather than $-r/2$ first. From there, the expected time to get to either 0 or r is again $\mathbb{E}[t_{r/2}]$. At that point, we are either done (we have reached r), or we are back at 0 (with fifty percent chance). In the latter case, we need to restart the entire process: our expected time to get to $\pm r$ is once again $\mathbb{E}[t_r]$. Thus, we have the following recursion:

$$\mathbb{E}[t_r] = 2\mathbb{E}[t_{r/2}] + 0.5 \cdot \mathbb{E}[t_r].$$

This, in turn, implies (A.1), which completes the proof. \square

We can now prove Proposition A.1 with a simple application of Markov's inequality.

Proof. [Proof of Proposition A.1] First observe that, if there exists any $i \leq n$ for which $Z_i \geq \sqrt{n}/4$, then with probability at least $1/2$ we will also have $Z_n \geq \sqrt{n}/4$. This is because the portion of the random walk determined by X_{i+1}, \dots, X_n has (by symmetry) at least a $1/2$ chance of being non-negative. It follows that, to prove $\Pr[Z_n \geq \sqrt{n}/4] \geq 1/4$, it suffices to show that

$$\Pr[\exists i \leq n \text{ such that } Z_i \geq \sqrt{n}/4] \geq 1/2.$$

Let r be the power of two in the range $[\sqrt{n}/4, \sqrt{n}/2)$. If $t_r \leq n$, then there exists $i \leq n$ such that $Z_i \geq 0.1\sqrt{n}$. It therefore suffices to show that

$$\Pr[t_r \leq n] \geq 1/2.$$

By Claim A.1, we have that $\mathbb{E}[t_r] = r^2 \leq n/4$. It follows by Markov's Inequality that $\Pr[t_r \geq n] \leq 1/4$, and therefore that $\Pr[t_r \leq n] \geq 3/4 \geq 1/2$, as desired. \square

³Indeed, one can argue that every r steps, there is a probability of at least $1/2^r$ that we escape the interval $(-r, r)$. It follows that, after kr steps, the probability that we fail to escape the interval is $(1 - 1/2^r)^k$, which goes to 0 as k goes to infinity.