Abstract

We give upper concentration bounds for martingales that are uniform over finite times and extend classical Hoeffding and Bernstein bounds. They shed light on the relationship between the central limit theorem and the law of the iterated logarithm in finite time. We also prove our upper bounds to be optimal in a strong sense with a lower bound.

1 Introduction

Martingales are indispensable in studying the temporal dynamics of stochastic processes arising in a multitude of fields [7, 11]. Particularly when such processes have complex long-range dependences, it is often of interest to concentrate martingales uniformly over time.

On the theoretical side, a fundamental limit to such concentration is expressed by the law of the iterated logarithm (LIL). However, this only concerns asymptotic behavior, and many applications instead require a concentration result that holds uniformly over all finite times.

This manuscript presents such bounds for the large classes of martingales which are addressed by Hoeffding [8] and Bernstein [6] inequalities. The new results are essentially optimal, and can be viewed as finite-time versions of the upper half of the LIL.

To be concrete, the simplest nontrivial martingale for such purposes is the discrete-time random walk \( \{M_t\}_{t=0,1,2,...} \) induced by flipping a fair coin repeatedly. It can be written as \( M_t = \sum_{i=1}^{t} \sigma_i \), where \( \sigma_i \) are i.i.d. Rademacher-distributed random variables (Pr (\( \sigma_i = -1 \)) = Pr (\( \sigma_i = +1 \)) = 1/2), so we refer to it as the "Rademacher random walk"; take \( M_0 = 0 \) w.l.o.g.

The LIL was first discovered for the Rademacher random walk, by Khinchin:

**Theorem 1 (Law of the iterated logarithm [10]).** Suppose \( M_t \) is a Rademacher random walk. Then with probability 1,

\[
\limsup_{t \to \infty} \frac{|M_t|}{\sqrt{t \log \log t}} = \sqrt{2}
\]

Our main result for the Rademacher random walk generalizes this to hold over finite times.

**Theorem 2.** Suppose \( M_t \) is a Rademacher random walk. Then there is an absolute constant \( C = 173 \) such that for any \( \delta < 1 \), with probability \( \geq 1 - \delta \), for all \( t \geq C \log(\frac{4}{\delta}) \) simultaneously, the following are true: \( |M_t| \leq \frac{t}{e^2(1+\sqrt{1/3})} \) and \( |M_t| \leq \sqrt{3t \left( 2 \log \log \left( \frac{4t}{2|M_t|} \right) \right) + \log \left( \frac{4}{\delta} \right)} \). The latter implies \( |M_t| \leq \max \left( \sqrt{3t \left( 2 \log \log \left( \frac{4t}{2|M_t|} \right) + \log \left( \frac{4}{\delta} \right) \right)}, 1 \right) \).
Theorem 2 takes the form of the LIL upper bound as \( t \to \infty \) for any fixed \( \delta > 0 \). Interestingly, it also captures a finite-time tradeoff between \( t \) and \( \delta \). The \( \log \log t \) term is independent of \( \delta \), and is therefore dominated by the \( \log \left( \frac{1}{\delta} \right) \) term for \( t \lesssim e^{1/\delta} \). In this regime, the bound is \( O \left( \sqrt{t \log \left( \frac{1}{\delta} \right)} \right) \) uniformly over time, a uniform central limit theorem (CLT)-type bound below the LIL rate for finite time and small \( \delta \). This is of applied interest because \( e^{1/\delta} \) can often be extremely large, in which case the CLT regime can encompass all times realistically encountered in practice.

1.1 Optimality of Theorem 2

We now show that Theorem 2 is optimal in a very strong sense. Suppose we are concerned with concentration of the random walk uniformly over time up to some fixed finite time \( T \). If the failure probability \( \delta \lesssim \frac{1}{\log T} \), then the \( \log \frac{1}{\delta} \) term dominates the \( \log \log t \) term for all \( t < T \). In this case, the bound of Theorem 2 is \( O \left( \sqrt{t \log \frac{1}{\delta}} \right) \) uniformly over \( t < T \). This is optimal even for a fixed \( t \) by binomial tail lower bounds, e.g. Slud’s inequality.

The more interesting case for our purposes is when \( \delta \gtrsim \frac{1}{\log T} \), in which case Theorem 2 gives a concentration rate of \( O \left( \sqrt{t \log \log t} + t \log \frac{1}{\delta} \right) \), uniformly over \( t < T \). As \( T \) and \( t \) increase without bound for any fixed \( \delta > 0 \), this rate becomes \( O \left( \sqrt{t \log \log t} \right) \), which is unimprovable by the LIL.

But the tradeoff between \( t \) and \( \delta \) given in Theorem 2 is also optimal up to small constants, as shown by the following result.

**Theorem 3.** There are absolute constants \( C_1, C_2 \) such that the following is true. Fix a finite time \( T > C_2 \log \left( \frac{T}{\delta} \right) \), and fix any \( \delta \in \left[ \frac{4}{\log((T-1)/3)}, \frac{1}{C_1} \right] \). Then with probability \( \leq 1 - \delta \), for all \( t \in \left[ C_2 \log \left( \frac{T}{\delta} \right), T \right) \) simultaneously, \( |M_t| \leq \frac{2}{3} t \) and

\[
|M_t| \leq \sqrt{\frac{2}{3} t \left( \log \log \left( \frac{T}{\delta} \right) + 2 \sqrt{T/3} \right) + \log \left( \frac{1}{C_1 \delta} \right)}
\]

(It suffices if \( C_1 = \left( \frac{420}{11} \right)^2 \) and \( C_2 = 164 \).)

We know of no previous results in this vein, so a principal contribution of this manuscript is to characterize the tradeoff between \( t \) and \( \delta \) in uniform concentration of measure over time. Theorems 2 and 3 constitute a tight finite-time version of the LIL, in the same way that the Hoeffding bound for large deviations is a tight finite-time version of the CLT’s Gaussian tail for a fixed time.

The proofs of both results are possibly of independent interest. The proof of Theorem 2 extends the exponential moment method, the standard way of proving classical Chernoff-style bounds which hold for a fixed time. We use a technique, manipulating stopping times of a particular averaged supermartingale, which generalizes easily to many discrete- and continuous-time martingales. These martingale generalizations are given in Section 2 with proof details in Section 3.

The proof of Theorem 3 basically inverts the argument used to prove the upper bound of Theorem 2 so we relegate it to Appendix E.

2 Uniform Upper Bounds for General Martingales

In this section, we extend the random walk concentration result of Theorem 2 to broader classes of martingales. Some notation must be established first.
We study the behavior of a real-valued stochastic process \( M_t \) in a filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)\), where \( M_0 = 0 \) w.l.o.g. For simplicity, only the discrete-time case \( t \in \mathbb{N} \) is considered hereafter; the results and proofs in this manuscript extend to continuous time as well.

Define the difference sequence \( \xi_t = M_t - M_{t-1} \) for all \( t \) (note that \( \xi_t \) is \( \mathcal{F}_t \)-measurable), and the cumulative conditional variance and quadratic variation: \( V_t = \sum_{i=1}^t \mathbb{E} [\xi_i^2 \mid \mathcal{F}_{i-1}] \) and \( Q_t = \sum_{i=1}^t \xi_i^2 \) respectively.

Also recall the following standard definitions. A martingale \( M_t \) (resp. supermartingale, submartingale) has \( \mathbb{E} [\xi_t \mid \mathcal{F}_{t-1}] = 0 \) (resp. \( \leq 0 \), \( \geq 0 \)) for all \( t \). A stopping time \( \tau \) is a function on \( \Omega \) such that \( \{\tau \leq t\} \in \mathcal{F}_t \ \forall t; \) notably, \( \tau \) can be infinite with positive probability \( [5] \).

### 2.1 Uniform Bernstein-Type Martingale Concentration

A few pertinent generalizations of Theorem 2 are now presented. The first is a direct uniform analogue of Bernstein’s inequality for martingales.

**Theorem 4** (Uniform Bernstein Bound). Let \( M_t \) be a martingale. Suppose the difference sequence is uniformly bounded \( [4] \); \( |M_t - M_{t-1}| \leq e^2 \) a.s. \( 1 \) for all \( t \geq 1 \). Fix any \( \delta < 1 \) and define \( \tau_0 = \min \{ \tau : 2(e-2)V_s \geq 173 \log \left( \frac{e}{\delta} \right) \} \). Then with probability \( \geq 1 - \delta \), for all \( t \geq \tau_0 \) simultaneously, \( |M_t| \leq \frac{2(e-2)}{e^2(1+\sqrt{1/3})} V_t \) and

\[
|M_t| \leq \sqrt{6(e-2)V_t \left( 2 \log \log \left( \frac{3(e-2)V_t}{|M_t|} \right) + \log \left( \frac{2}{\delta} \right) \right)}
\]

Theorem 4 is particularly convenient for many applications because the cumulative conditional variance \( V_t \) marginalizes over the present time \( t \), and therefore can often be controlled usefully in practice.

Another generalization of Theorem 2 is as follows.

**Theorem 5.** Let \( M_t \) be a martingale. Fix any \( \delta < 1 \) and define \( \tau_0 = \min \{ \tau : \frac{1}{2}(2V_s + Q_s) \geq 173 \log \left( \frac{e}{\delta} \right) \} \). Then with probability \( \geq 1 - \delta \), for all \( t \geq \tau_0 \) simultaneously, \( |M_t| \leq \frac{2V_t + Q_t}{3e^2(1+\sqrt{1/3})} \) and

\[
|M_t| \leq \sqrt{(2V_t + Q_t) \left( 2 \log \log \left( \frac{2V_t + Q_t}{2|M_t|} \right) + \log \left( \frac{2}{\delta} \right) \right)}
\]

This result does not impose any conditions at all on \( \xi_i \), and involves \( Q_t \) in order to avoid such requirements. But if each difference iterate is assumed a.s. bounded in Theorem 5, a uniform counterpart to the Hoeffding inequality is the direct corollary.

**Theorem 6** (Uniform Hoeffding Bound). Let \( M_t \) be a martingale, and suppose there are constants \( \{c_i\}_{i \geq 1} \) such that for all \( t \geq 1 \), \( |M_t - M_{t-1}| \leq c_t \) a.s. \( 1 \). Fix any \( \delta < 1 \) and define \( \tau_0 = \min \{ \tau : \sum_{i=1}^t c_i^2 \geq 173 \log \left( \frac{e}{\delta} \right) \} \). Then with probability \( \geq 1 - \delta \), for all \( t \geq \tau_0 \) simultaneously,

\[
|M_t| \leq \left( \frac{1}{e^2(1+\sqrt{1/3})} \right) \left( \sum_{i=1}^t c_i^2 \right) \text{ and}
|M_t| \leq \sqrt{3 \left( \sum_{i=1}^t c_i^2 \right) \left( 2 \log \log \left( \frac{3 \left( \sum_{i=1}^t c_i^2 \right)}{2|M_t|} \right) + \log \left( \frac{2}{\delta} \right) \right)}
\]

\(^1\)As with any Bernstein-type inequality, the boundedness assumption on \( \xi_t \) can be replaced by higher moment conditions.

\(^2\)Throughout we use the convention that the minimum and maximum of an empty set are \( \infty \) and \( -\infty \) respectively.
The proofs of Theorems 4 and 5 are nearly identical to that of Theorem 2, further details are given in Appendix B.

2.2 Extensions and Remarks

Extension to Continuous-time Martingales. In many cases, these uniform results can be generalized to continuous-time martingales with an almost unchanged proof (e.g., this is true for the Wiener process \( W_t \)). Further explanation of this depends on the proof details, and therefore is deferred to Section 3.3. □

Super(sub)martingale Bounds. One-sided variants of Theorem 2 hold in many cases for super-(resp. sub-) martingales, giving a uniform upper (resp. lower) bound identical to that in Theorem 2. When the Doob-Meyer decomposition (5) applies, as is frequently the case, such bounds are immediate. □

Other Maximal Bounds. When considering uniform martingale concentration over all times without an explicit union bound, the basic tools are Doob’s maximal inequality for nonnegative supermartingales (5, Exercise 5.7.1), Hoeffding’s maximal inequality [8], and Freedman’s Bernstein-type inequality [6]. These can all be easily proved with the techniques of this manuscript (similar to the proof of Theorem 11). However, the latter two results are fundamentally weaker than ours, in that they only hold uniformly over a finite time interval, and degrade to triviality as the interval grows infinite. □

Initial Time Conditions. The upper bounds in this manuscript include a condition \( t \geq \tau_0 \) for \( \tau_0 = \min \{ s : U_s = 173 \log \left( \frac{1}{\delta} \right) \} \), where \( U_t \) is a nondecreasing process measuring cumulative variance. For Theorem 2, it is straightforward to remove this condition without degrading the result: if \( t < \tau_0 \), an explicit union bound over fixed-time Hoeffding bounds immediately gives that \( |M_t| \leq O \left( \log \left( \frac{1}{\delta} \right) \right) \) w.h.p. for all \( t < \tau_0 \). Combining this with Theorem 2 gives a uniform bound, over all times \( t \geq 0 \), of \( |M_t| \leq O \left( \sqrt{t} \left( \log \log t + \log \log \frac{1}{\delta} \right) + \log \left( \frac{1}{\delta} \right) \right) \). The last \( \log \left( \frac{1}{\delta} \right) \) term exactly matches standard Bernstein/Bennett concentration inequalities for a fixed time.

To extend this argument to the upper bounds of Theorems 4-6, such an explicit union bound is no longer usable. But our proof techniques using stopping times extend naturally to these cases - Appendix C contains the details. □

Tightening the Bounds. The leading proportionality constant on the \( \sqrt{t \log \log t} \) term in Theorem 2 is \( \sqrt{6} \), clearly suboptimal in the limit \( t \to \infty \) by the LIL. This constant can be lowered arbitrarily close to optimality as \( t \) increases. This suggests that our proof technique is quite tight.

The tightened proofs may be of interest because they generalize results of Robbins and Siegmund [12, 13] to the non-asymptotic case, and to general martingales. Appendix D contains further details. □

2.3 Discussion

Our uniform Hoeffding/Bernstein-type bounds in Section 2.1 achieve optimal rates in the variance and \( \delta \) parameters as well. These generalize martingale LILs like the classic result of Stout [15], which for large classes of martingales makes a statement similar to Theorem 1, except concerning the ratio \( \frac{|M_t|}{\sqrt{V_t \log \log V_t}} \).

\footnote{After this work was completed, the author became aware of another recent finite-time upper bound LIL, restricted to i.i.d. sub-Gaussian difference sequences [9]. It is proved with an epoch-based approach standard in proofs of the (asymptotic) LIL. This manuscript can be viewed as generalizing that idea using stopping time manipulations.}
Theorem 2's tradeoff between $t$ and $\delta$ describes some of the interplay between the CLT and the LIL when uniform bounds are taken of partial sums of suitable i.i.d. variables. The same question has been explored with a different statistical emphasis by Darling and Erdős [1] and subsequent work, though only as $t \to \infty$ to our knowledge.

3 Proving Theorem 2

Define the (deterministic) process $U_t = t.$

Also define $k := \frac{1}{3}$ and $\lambda_0 := \frac{1}{e^{2(1+\sqrt{k})}}$ for this section.

In this section, we prove the following bound, which is a slight more precise version of Theorem 2:

**Theorem 7.** Let $M_t$ be a Rademacher random walk. Fix any $\delta < 1$ and define the time $\tau_0 = \min \{ s : U_s \geq \frac{2}{\lambda_0} \log \left( \frac{4}{\delta} \right) \}$. Then with probability $\geq 1 - \delta$, for all $t \geq \tau_0$ simultaneously, $|M_t| \leq \lambda_0 U_t$ and

$$|M_t| \leq \frac{2U_t}{1 - k \log \left( \frac{2 \log \left( \frac{U_t}{(1-k)|M_t|} \right) \delta}{\delta} \right)}.$$

The proof invokes the Optional Stopping Theorem in martingale theory, in particular a version for nonnegative supermartingales that exploits their favorable convergence properties:

**Theorem 8** (Optional Stopping for Nonnegative Supermartingales ([5], Theorem 5.7.6)). Let $M_t$ be a nonnegative supermartingale. Then if $\tau$ is a (possibly infinite) stopping time, $\mathbb{E}[M_\tau] \leq \mathbb{E}[M_0]$.

The argument begins by appealing to a standard exponential supermartingale construction.

**Lemma 9.** The process $X_\lambda^t := \exp \left( \lambda M_t - \frac{\lambda^2}{2} U_t \right)$ is a supermartingale for any $\lambda \in \mathbb{R}$.

**Proof.** Using Hoeffding’s Lemma, for any $\lambda \in \mathbb{R}$ and $t \geq 1$, $\mathbb{E} [\exp (\lambda \xi_t) | F_{t-1}] \leq \exp \left( \frac{\lambda^2}{2} (2^2) \right) = \exp \left( \frac{\lambda^2}{2} \right)$. Therefore, $\mathbb{E} \left[ \exp \left( \lambda \xi_t - \frac{\lambda^2}{2} \right) | F_{t-1} \right] \leq 1$, so $\mathbb{E} \left[ X_\lambda^t | F_{t-1} \right] \leq X_\lambda^{t-1}$. ■

The result is derived through various manipulations of this supermartingale $X_\lambda^t$.

For the rest of the proof, for all $t$, assume that $M_t \neq 0$. This is without loss of generality, because when $M_t = 0$, the bound of Theorem 2 trivially holds.

3.1 A Time-Uniform Law of Large Numbers

The desired result, Theorem 7, uniformly controls $\frac{|M_t|}{\sqrt{U_t \log \log U_t}}$, but we first control $\frac{|M_t|}{U_t}$. This generalizes the (strong) law of large numbers ([S]LLN), for any failure probability $\delta > 0$, uniformly over finite times.

While a weaker result than Theorem 7, this concisely demonstrates our principal proof techniques, and is independently necessary as a “bootstrap” for the main bound.

The first step is to establish a moment bound which holds at any stopping time, by averaging supermartingales from the uncountable family $\left\{ \exp \left( \lambda M_t - \frac{\lambda^2}{2} U_t \right) \right\}_{\lambda \in \mathbb{R}}$ using a particular weighting over $\lambda$.

---

4 A notational convenience, to ease extension of this proof to the martingale case discussed in Section 2.
Lemma 10. For any stopping time \( \tau \), \( \mathbb{E} \left[ \exp \left( \lambda_0 |M\tau| - \frac{\lambda_0^2}{2} U\tau \right) \right] \leq 2. \)

**Proof.** Recall the definition of \( X_\lambda \) from Lemma 9. Here we set the free parameter \( \lambda \) in the process \( X_\lambda \) to get a process \( Y_t \). \( \lambda \) is set stochastically: \( \lambda \in \{-\lambda_0, \lambda_0\} \) with probability \( \frac{1}{2} \) each. After marginalizing over \( \lambda \), the resulting process is

\[
Y_t = \frac{1}{2} \exp \left( \lambda_0 M_t - \frac{\lambda_0^2}{2} U_t \right) + \frac{1}{2} \exp \left( -\lambda_0 M_t - \frac{\lambda_0^2}{2} U_t \right) \geq \frac{1}{2} \exp \left( \lambda_0 |M_t| - \frac{\lambda_0^2}{2} U_t \right)
\]

(1)

Now take \( \tau \) to be any stopping time as in the lemma statement. Then \( \mathbb{E} \left[ \exp \left( \lambda_0 M_\tau - \frac{\lambda_0^2}{2} U_\tau \right) \right] = \mathbb{E} \left[ X_\lambda = \lambda_0 \right] \leq 1 \), where the inequality is by the Optional Stopping Theorem (Theorem 8). Similarly, \( \mathbb{E} \left[ X_\lambda = -\lambda_0 \right] \leq 1 \).

So \( \mathbb{E} [Y_\tau] = \frac{1}{2} (\mathbb{E} \left[ X_\lambda = \lambda_0 \right] + \mathbb{E} \left[ X_\lambda = -\lambda_0 \right]) \leq 1 \). Combining this with (1) gives the result. \( \blacksquare \)

A particular setting of \( \tau \) extracts the desired uniform LLN bound from Lemma 10.

**Theorem 11.** Fix any \( \delta > 0 \). With probability \( \geq 1 - \delta \), for all \( t \geq \min \left\{ t : U_t \geq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right\} \) simultaneously, \( \frac{|M|}{U_t} \leq \lambda_0 \)

**Proof.** For convenience define \( \tau_1 = \min \left\{ t : U_t \geq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right\} \). Define the stopping time \( \tau = \min \left\{ t \geq \tau_1 : \frac{|M|}{U_t} > \lambda_0 \right\} \). Then it suffices to prove that \( P(\tau < \infty) \leq \delta \).

On the event \( \{ \tau < \infty \} \), we have \( \frac{|M|}{U_t} \leq \lambda_0 \) by definition of \( \tau \). Therefore, using Lemma 10

\[
2 \geq \mathbb{E} \left[ \exp \left( \lambda_0 |M\tau| - \frac{\lambda_0^2}{2} U\tau \right) \right] \geq \mathbb{E} \left[ \exp \left( \lambda_0 |M\tau| - \frac{\lambda_0^2}{2} U\tau \right) \mid \tau < \infty \right] P(\tau < \infty)
\]

\[
\geq \mathbb{E} \left[ \exp \left( \frac{\lambda_0^2}{2} U\tau - \frac{\lambda_0^2}{2} U\tau \right) \right] P(\tau < \infty) = \mathbb{E} \left[ \exp \left( \frac{\lambda_0^2}{2} U\tau \right) \right] P(\tau < \infty) \geq \frac{2}{\delta} P(\tau < \infty)
\]

where \( (a) \) uses that \( \frac{|M|}{U_t} > \lambda_0 \) when \( \tau < \infty \), and \( (b) \) uses \( U \geq \tau_1 \geq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \). Therefore, \( P(\tau < \infty) \leq \delta \), as desired. \( \blacksquare \)

The process \( U_t \) is increasing in any case of interest, implying that \( \frac{|M|}{U_t} \leq \lambda_0 \) uniformly in \( t \) after some finite initial time. The setting of \( \lambda_0 \) happens to fit with the rest of our LIL proof, but this choice of \( \lambda_0 \) in Theorem 11 is arbitrary. The same proof method in fact defines a family of bounds parametrized by \( \lambda_0 \); collectively, these can be considered to generalize the SLLN.

### 3.2 Main Proof

We proceed to prove Theorem 7 using the SLLN bound of Theorem 11 and its proof techniques.

#### 3.2.1 Preliminaries

A little further notation is required for the rest of the proof.

For any event \( E \subseteq \Omega \) of nonzero measure, let \( \mathbb{E}_E [\cdot] \) denote the expectation restricted to \( E \), i.e. \( \mathbb{E}_E [f] = \frac{1}{|E|} \int_E f(\omega) P(d\omega) \) for a measurable function \( f \) on \( \Omega \). Similarly, dub the associated measure \( P_E \), where for any event \( \Xi \subseteq \Omega \) we have \( P_E(\Xi) = \frac{P(E \cap \Xi)}{P(E)}. \)
Consider the “good” event of Theorem 11 in which its uniform deviation bound holds w.p. $\geq 1 - \delta$ for some $\delta$; call this event $A_\delta$. Formally,

$$A_\delta = \{ \omega \in \Omega : \frac{|M_t|}{U_t} \leq \lambda_0 \, \forall t \geq \min \left\{ s : U_s \geq 2^{\lambda_0^2 \log \left( \frac{2}{\delta} \right)} \right\} \}$$

(2)

Theorem 11 states that $P(A_\delta) \geq 1 - \delta$.

It will be necessary to shift sample spaces from $A_\delta$ to $\Omega$. The shift should be small in measure because $P(A_\delta) \geq 1 - \delta$; this is captured by the following simple observation.

**Lemma 12.** Define $A_\delta$ as in (2). For any nonnegative random variable $X$ on $\Omega$,

$$\mathbb{E}[X] \leq \frac{1}{1 - \delta} \mathbb{E}[X]$$

*Proof.* Since $X \geq 0$, using Thm. 11, 

$$\mathbb{E}[X] = \mathbb{E}_{A_\delta}[X] P(A_\delta) + \mathbb{E}_{A_\delta^c}[X] P(A_\delta^c) \geq \mathbb{E}_{A_\delta}[X] (1 - \delta).$$

3.2.2 Proof of Theorem 7

The main result can now be proved. The first step is to choose $\lambda$ stochastically in the supermartingale $X_\lambda$ and bound the effect of averaging over $\lambda$ (analogous to Lemma 10 in the proof of the bootstrap bound).

**Lemma 13.** Define $\tau_0$ as in Theorem 7 and $A_\delta$ as in (2) for any $\delta$. Then for any stopping time $\tau \geq \tau_0$,

$$\mathbb{E}_{A_\delta} \left[ \mathbb{E}^\lambda \left[ X^\lambda_\tau \right] \right] \geq \mathbb{E}_{A_\delta} \left[ \frac{2 \exp \left( \frac{M^2_\tau}{2U_\tau} (1 - k) \right)}{\log^2 \left( \frac{U_\tau}{(1 - \sqrt{k}) |M_\tau|} \right)} \right]$$

(3)

The proof relies on estimating an integral and is deferred to Appendix A.

Lemma 13 can be converted into the desired uniform bound using a particular choice of stopping time, analogously to how the bootstrap bound Theorem 11 is derived from Lemma 10. However, this time a shift in sample spaces is also needed to yield Theorem 7 since Lemma 13 uses $A_\delta$ instead of $\Omega$.

*Proof of Theorem 7.* Define the stopping time

$$\tau = \min \left\{ t \geq \tau_0 : |M_t| > \lambda_0 U_t \lor \begin{cases} |M_t| \leq \lambda_0 U_t \land |M_t| > \frac{2U_t}{1 - k} \log \left( \frac{2 \log^2 \left( \frac{U_t}{(1 - \sqrt{k}) |M_t|} \right)}{\delta} \right) \end{cases} \right\}$$

(3)
It suffices to prove that $P(\tau = \infty) \geq 1 - \delta$. On the event $\{\tau < \infty\} \cap A_{\delta/2}$, we have

$$|M_{\tau}| > \sqrt{\frac{2U_{\tau}}{1 - k} \log \left( \frac{2\log^2 \left( \frac{U_{\tau}}{1 - \sqrt{k}|M_{\tau}|} \right)}{\delta} \right)} \iff 2 \exp \left( \frac{M^2_{\tau}}{2U_{\tau}} (1 - k) \right) > \frac{4}{\delta}$$

Therefore, using Lemma 13 and the nonnegativity of $\frac{2 \exp \left( \frac{M^2_{\tau}}{2U_{\tau}} (1 - k) \right)}{\log^2 \left( \frac{U_{\tau}}{1 - \sqrt{k}|M_{\tau}|} \right)}$ on $A_{\delta/2}$,

$$2 \geq \frac{1}{1 - \frac{\delta}{2}} = \mathbb{E}^\lambda \left[ \mathbb{E} \left[ X_\delta^\lambda \right] \right] \geq \frac{\mathbb{E}^\lambda \left[ \mathbb{E} \left[ X_k^\lambda \right] \right]}{1 - \frac{\delta}{2}} \geq \mathbb{E}^\lambda \left[ \mathbb{E}_{A_{\delta/2}} \left[ X_\tau^\lambda \right] \right] \geq \mathbb{E}_{A_{\delta/2}} \left[ \frac{2 \exp \left( \frac{M^2_{\tau}}{2U_{\tau}} (1 - k) \right)}{\log^2 \left( \frac{U_{\tau}}{1 - \sqrt{k}|M_{\tau}|} \right)} \right] \geq \mathbb{E}_{A_{\delta/2}} \left[ \mathbb{E} \left[ X_\tau^\lambda \right] \right] \geq \frac{4}{\delta} P_{A_{\delta/2}}(\tau < \infty)$$

where (a) is by Optional Stopping (Theorem 8; note that $\tau$ can be unbounded), (b) is by Lemma 12, (c) is by Tonelli’s Theorem, (d) is by Lemma 13 and (e) is by the definitions of $A_{\delta/2}$ in (2) and of $\tau$ in (3).

After simplification, this gives

$$P_{A_{\delta/2}}(\tau < \infty) \leq \delta/2 \implies P_{A_{\delta/2}}(\tau = \infty) \geq 1 - \frac{\delta}{2} \quad (4)$$

Therefore,

$$P(\tau = \infty) \geq P(\{\tau = \infty\} \cap A_{\delta/2}) \geq P_{A_{\delta/2}}(\tau = \infty) P(A_{\delta/2}) \geq \left( 1 - \frac{\delta}{2} \right) \left( 1 - \frac{\delta}{2} \right) \geq 1 - \delta$$

where (a) uses the definition of $P_{A_{\delta/2}}(\cdot)$ and (b) uses (4) and Thm. 11. The result follows.

### 3.3 Proof Discussion

Most of the tools used in this proof, particularly optional stopping as in Theorem 8, extend seamlessly to the continuous-time case. The only potential obstacle to this is in the first step - establishing an exponential supermartingale construction of the form of Lemma 9. This is easily done in many situations of interest, as demonstrated by the archetypal result that the standard geometric Brownian motion $\exp \left( \lambda W_t - \frac{\lambda^2}{2} t \right)$ is a martingale for any $\lambda \in \mathbb{R}$, where $W_t$ is the standard Wiener process. Indeed, the exponential construction is tight in this case, unlike in discrete time where it is merely a supermartingale.

The proof of this manuscript is possible because the index set (time) is totally ordered, and can be manipulated using filtrations and stopping times. There is a very interesting analogy to well-developed general chaining techniques 10 that have been used to great effect to uniformly bound processes indexed on metric spaces, by using covering arguments which incorporate variation at different scales (12, e.g. Problem 12.14). Exploration of such relationships is left open.
A direct antecedent to this manuscript is a pioneering line of work by Robbins and colleagues [2, 12, 13] that investigates the powerful method of averaging martingales. For the most part, it only considers the asymptotic regime, though Darling and Robbins [2] do briefly treat finite times (with far weaker $\delta$ dependence). More recently, de la Peña et al. [3] revisit their techniques with a different emphasis and normalization for $M_t$.

The idea of using stopping times in the context of uniform martingale concentration goes back at least to work of Robbins [13] with Siegmund [12], and was then notably used by Freedman [6].

Our proof techniques are conceptually related to ideas from Shafer and Vovk ([14], Ch. 5), who describe how to view the LIL as emerging from a game. Departing from traditional approaches, they motivate the exponential supermartingale construction and prove the (asymptotic) LIL by averaging such supermartingales.

Acknowledgements

I am very grateful to Sanjoy Dasgupta, Patrick Fitzsimmons, Yoav Freund, and Matus Telgarsky for instructive conversations.

---

5This can otherwise be motivated with the continuous-time case, where it is an exact martingale due to CLT effects in the Donsker continuous-time limit. The book [14] gives a somewhat different argument using Taylor expansions.
References

[1] D. A. Darling and P. Erdős. A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal*, 23(1):143–155, 1956.

[2] D. A. Darling and H. Robbins. Iterated logarithm inequalities. *Proc. Nat. Acad. Sci. U.S.A.*, 57:1188–1192, 1967.

[3] Victor H. de la Peña, Michael J. Klass, and Tze Leung Lai. Pseudo-maximization and self-normalized processes. *Probability Surveys*, 4:172–192, 2007.

[4] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer, 1996.

[5] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4th edition, 2010.

[6] David A. Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

[7] C. C. Heyde. Martingales: A case for a place in the statistician’s repertoire. *Australian Journal of Statistics*, 14(1):1–9, 1972.

[8] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[9] Kevin G. Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil’ ucb : An optimal exploration algorithm for multi-armed bandits. In Maria-Florina Balcan and Csaba Szepesvári, editors, *COLT*, volume 35 of *JMLR Proceedings*, pages 423–439. JMLR.org, 2014.

[10] A. Ya. Khinchin. über einen satz der wahrscheinlichkeitsrechnung. *Fundamenta Mathematicae*, 6:9–20, 1924.

[11] Yuval Peres. The unreasonable effectiveness of martingales. In Claire Mathieu, editor, *SODA*, pages 997–1000. SIAM, 2009.

[12] H. Robbins and D. Siegmund. Boundary crossing probabilities for the wiener process and sample sums. *The Annals of Mathematical Statistics*, 41:1410–1429, 1970.

[13] Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, October 1970.

[14] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It’s Only a Game!* Wiley, 1st edition, 2001.

[15] William F. Stout. A martingale analogue of kolmogorov’s law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete*, 15:279–290, 1970.

[16] Michel Talagrand. *The Generic Chaining: Upper and Lower Bounds for Stochastic Processes*. Springer, 1st edition, 2005.
A Proof of Lemma [13]

Proof of Lemma [13] As outlined in Section 3.2.1 we choose \( \lambda \) stochastically from a probability space \( (\Omega, \mathcal{F}, P_\lambda) \) such that \( P_\lambda(d\lambda) = \frac{d\lambda}{|\lambda| \left( \log \frac{1}{|\lambda|} \right)^2} \) on \( \lambda \in [-e^{-2}, e^{-2}] \setminus \{0\} \).

Take an arbitrary time \( t \geq \tau_0 \). For outcomes within \( A_\delta \), we have the following:

\[
\mathbb{E}_\lambda^X \left[ X_t^\lambda \right] = \int_{-1/e^2}^{1/e^2} \exp \left( \lambda M_t - \frac{\lambda^2}{2} U_t \right) \frac{d\lambda}{-\left( \lambda \left( \log \frac{1}{\lambda} \right)^2 \right)} + \int_0^{1/e^2} \exp \left( \lambda M_t - \frac{\lambda^2}{2} U_t \right) \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
\geq \exp \left( \frac{M_t^2}{2U_t} \right) \int_{-1/e^2}^{1/e^2} e^{-1/2 U_t \left( \lambda - \frac{M_t}{U_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \int_0^{1/e^2} e^{-1/2 U_t \left( \lambda - \frac{M_t}{U_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
\geq \exp \left( \frac{M_t^2}{2U_t} \right) \int_0^{1/e^2} e^{-1/2 U_t \left( \lambda - \frac{M_t}{U_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
\geq \exp \left( \frac{M_t^2}{2U_t} \right) \exp \left( -\frac{1}{2} k U_t \left( \frac{M_t}{U_t} \right)^2 \right) \int_0^{1/e^2} e^{-1/2 U_t \left( \lambda - \frac{M_t}{U_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
= \exp \left( \frac{M_t^2}{2U_t} (1 - k) \right) \frac{\log \left( \frac{U_t}{|M_t| (1 + \sqrt{k})} \right)}{\log \left( \frac{U_t}{|M_t| (1 - \sqrt{k})} \right)} \\
\geq 2 \exp \left( \frac{M_t^2}{2U_t} (1 - k) \right) \frac{1}{\log \left( \frac{U_t}{|M_t| (1 + \sqrt{k})} \right)} \frac{\log \left( \frac{U_t}{|M_t| (1 - \sqrt{k})} \right)}{\log \left( \frac{U_t}{|M_t| (1 - \sqrt{k})} \right)} \geq \frac{2 \exp \left( \frac{M_t^2}{2U_t} (1 - k) \right)}{\log^2 \left( \frac{U_t}{|M_t| (1 - \sqrt{k})} \right)}
\]

Take \( \tau \) to be any stopping time as in the lemma statement. Then from (6),

\[
\mathbb{E}_{A_\delta} \mathbb{E}_\lambda^X \left[ X_{\tau}^\lambda \right] \geq \mathbb{E}_{A_\delta} \left[ \frac{2 \exp \left( \frac{M_t^2}{2U_t} (1 - k) \right)}{\log^2 \left( \frac{U_t}{|M_t| (1 - \sqrt{k})} \right)} \right]
\]

B Generalizations of Theorem [2]

In this section, the results of Section [2] are justified, by appealing to the fact that they can be proved through simple extensions of the proof of Theorem [2].

That proof is the subject of Section [3]. It applies just to the Rademacher random walk, but uses the i.i.d. Rademacher assumption only through an exponential supermartingale construction (Lemma [4]). Theorem [2] can be generalized significantly beyond the Rademacher random walk by simply replacing the construction with other similar exponential constructions, leaving the remainder of the proof essentially intact as presented in Section [3]
To be specific, the rest of that proof works unchanged if the construction has the following properties:

1. The construction should be of the same form as Lemma \[9\] \( X^\lambda_t = \exp \left( \lambda M_t - \frac{\lambda^2}{2} U_t \right) \) for some nondecreasing process \( U_t \). (The proof of Theorem \[2\] sets \( U_t = t \).)

2. \( X^\lambda_t \) should be a supermartingale for \( \lambda \in \left( -\frac{1}{e^2}, \frac{1}{e^2} \right) \setminus \{0\} \).

Now we give two exponential supermartingale constructions with these properties. The first and second constructions lead directly to Theorems \[4\] and \[5\] respectively, when used to replace Lemma \[9\] in the proof of Theorem \[2\]. The first is standard, but the second may be of interest due to its lack of higher moment assumptions on \( \xi_t \).

**Lemma 14.** Suppose the difference sequence is uniformly bounded, i.e. \( |\xi_t| \leq e^2 \) a.s. for all \( t \). Then the process \( X^\lambda_t := \exp \left( \lambda M_t - \lambda^2 (e - 2)V_t \right) \) is a supermartingale for any \( \lambda \in \left[ -\frac{1}{e^2}, \frac{1}{e^2} \right] \).

**Proof.** It can be checked that \( e^x \leq 1 + x + (e - 2)x^2 \) for \( x \leq 1 \). Then for any \( \lambda \in \left[ -\frac{1}{e^2}, \frac{1}{e^2} \right] \) and \( t \geq 1 \),

\[
E \left[ \exp (\lambda \xi_t) \mid F_{t-1} \right] \leq 1 + \lambda E \left[ \xi_t \mid F_{t-1} \right] + \lambda^2 (e - 2) E \left[ \xi_t^2 \mid F_{t-1} \right]
\]

\[
= 1 + \lambda^2 (e - 2) E \left[ \xi_t^2 \mid F_{t-1} \right] \leq \exp \left( \lambda^2 (e - 2) E \left[ \xi_t^2 \mid F_{t-1} \right] \right)
\]

using the martingale property on \( E \left[ \xi_t \mid F_{t-1} \right] \).

Therefore, \( E \left[ \exp (\lambda \xi_t - \lambda^2 (e - 2)V_t) | F_{t-1} \right] \leq 1 \), so \( E \left[ X^\lambda_t \mid F_{t-1} \right] \leq X^\lambda_{t-1} \). \[\square\]

**Lemma 15.** The process \( X^\lambda_t := \exp \left( \lambda M_t - \frac{\lambda^2}{6} (2V_t + Q_t) \right) \) is a supermartingale for any \( \lambda \in \mathbb{R} \).

**Proof.** Consider the following inequality: for all real \( x \),

\[
\exp \left( x - \frac{1}{6} x^2 \right) \leq 1 + x + \frac{1}{3} x^2. \tag{7}
\]

Suppose (7) holds. Then for any \( \lambda \in \mathbb{R} \) and \( t \geq 1 \),

\[
E \left[ \exp \left( \lambda \xi_t - \frac{\lambda^2}{6} \xi_t^2 \right) \mid F_{t-1} \right] \leq 1 + \lambda E \left[ \xi_t \mid F_{t-1} \right] + \lambda^2 \frac{\lambda^2}{6} E \left[ \xi_t^2 \mid F_{t-1} \right] = 1 + \lambda^2 \frac{\lambda^2}{6} E \left[ \xi_t^2 \mid F_{t-1} \right] \leq \exp \left( \lambda^2 \frac{\lambda^2}{6} E \left[ \xi_t^2 \mid F_{t-1} \right] \right),
\]

using the martingale property on \( E \left[ \xi_t \mid F_{t-1} \right] \). Therefore, \( E \left[ \exp \left( \lambda \xi_t - \frac{\lambda^2}{6} \xi_t^2 \right) \mid F_{t-1} \right] \leq 1 \), so \( E \left[ X^\lambda_t \mid F_{t-1} \right] \leq X^\lambda_{t-1} \) and the result is shown.

It only remains to prove (7), which is equivalent to showing that the function \( f(x) = \exp \left( x - \frac{1}{6} x^2 \right) - 1 - x - \frac{1}{3} x^2 \) \( \leq 0 \). This is done by examining derivatives. Note that \( f'(x) = \left( 1 - \frac{x}{3} \right) \exp \left( x - \frac{1}{6} x^2 \right) - \frac{1}{2} - x - \frac{1}{3} x^2 \), \( f''(x) = \left( -\frac{1}{3} + \left( 1 - \frac{x}{3} \right)^2 \right) \exp \left( x - \frac{1}{6} x^2 \right) - \frac{2}{3} \) \( \leq 0 \) \( \exp (y(1-y)) \) when \( y := x - \frac{1}{6} x^2 \). Here \( e^y \leq \frac{1}{1-y} \) for \( y < 1 \), and \( e^{y}(1-y) \leq 0 \) for \( y \geq 1 \), so \( f''(x) \leq 0 \) for all \( x \). Since \( f'(0) = 0 = f(0) \), the function \( f \) attains a maximum of zero over its domain, proving (7) and the result. \[\square\]

\footnote{The constant \( \frac{1}{1-y} \) in these conditions is determined by the choice of averaging distribution over \( \lambda \) (i.e., \( P_\lambda \)) in the proof of Lemma \[13\]. See Appendix \[3\] for examples of other averaging distributions.}
C Initial Time Conditions

As discussed in Section 2.2 (under “Initial Time Conditions”), this section outlines how to remove the initial time condition for general martingales, using the proof technique of this manuscript. We follow the notation of Section 3 and extend Theorem 7 here, using the placeholder variance process \( U_t \) to handle the Hoeffding- and Bernstein-style results of Section 2. We can therefore use Lemmas 9 and 10.

It suffices to show a uniform concentration bound for \( |M_t| \) over \( t < \tau_1 := \min \left\{ s : U_s \geq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right\} \).

**Theorem 16.** Take any \( \delta > 0 \). With probability \( \geq 1 - \delta \), for all \( t < \tau_1 \) simultaneously,

\[
|M_t| \leq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right)
\]

**Proof.** Define the stopping time \( \tau = \min \left\{ t \leq \tau_1 : |M_t| > \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right\} \). Then it suffices to prove that \( P(\tau < \tau_1) \leq \delta \).

On the event \( \{ \tau < \tau_1 \} \), we have \( |M_\tau| > \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \) by definition of \( \tau \). Therefore, using Lemma 10,

\[
2 \geq \mathbb{E} \left[ \exp \left( \lambda_0 |M_\tau| - \frac{\lambda_0^2}{2} U_\tau \right) \right] \geq \mathbb{E} \left[ \exp \left( \lambda_0 |M_\tau| - \frac{\lambda_0^2}{2} U_\tau \right) \right] \mathbb{P}(\tau < \tau_1) \\
\geq \mathbb{E} \left[ \exp \left( \lambda_0 \left( \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right) - \frac{\lambda_0^2}{2} \left( \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \right) \right) \right] \mathbb{P}(\tau < \tau_1) \\
= \mathbb{E} \left[ \exp \left( \log \left( \frac{2}{\delta} \right) \right) \right] \mathbb{P}(\tau < \tau_1) = \frac{2}{\delta} \mathbb{P}(\tau < \tau_1)
\]

where (a) uses that \( |M_\tau| > \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \) when \( \tau < \tau_1 \), and that \( U_\tau \leq \frac{2}{\lambda_0} \log \left( \frac{2}{\delta} \right) \) since \( \tau < \tau_1 \). Therefore, \( P(\tau < \tau_1) \leq \delta \), as desired.

Taking a union bound of Theorem 16 with Theorem 7 gives that w.h.p., for all \( t \),

\[
|M_t| \leq O \left( \sqrt{U_t \left( \log \log U_t + \frac{1}{\delta} \right) + \log \left( \frac{1}{\delta} \right)} \right)
\]

This matches the rate of Bennett/Bernstein inequalities which hold for a fixed time, except for an extra \( \sqrt{\log \log U_t} \) factor on the Gaussian-regime term that accounts for the uniformity over time.

D Other Averaging Distributions for Sharper Constants

The leading proportionality constant on the iterated-logarithm term in Theorem 7 is \( \sqrt{6} \), above the LIL’s asymptotic \( \sqrt{2} \). The reasons for this relate to the proof of Lemma 13 (Appendix A).

First, the mixed process \( \mathbb{E}^\lambda \left[ X_\lambda^t \right] \) in this proof can be written as a probability integral of a Gaussian-like function, which we crudely lower-bound around the peak (Eq. 5). A more refined lower bound here would lead to a sharper final result. This accounts for a \( \sqrt{3} \) factor out of the \( \sqrt{6} \) leading constant. A better bound tightens this arbitrarily close to a (LIL-optimal) \( \sqrt{2} \) factor as \( t \to \infty \), because of the nature of the supermartingale construction.

For the rest of the current section, we neglect this source of looseness and only discuss how to lower the proportionality constant from \( \sqrt{6} \) arbitrarily close to \( \sqrt{3} \), by changing the averaging distribution for \( \lambda \). By the discussion above, this change is unimprovable.
D.1 A Family of Averaging Distributions

First, we present a countably infinite family of averaging distributions for \( \lambda \).

To describe this set, define \( \log_v(x) = \log \log \ldots \log (x) \) and \( \exp_v(x) = \exp \exp \ldots \exp (x) \) for \( v = 1, 2, \ldots \). The following family of probability distributions is indexed by \( v \):

\[
P_v^\lambda(d\lambda) = \frac{d\lambda}{|\lambda| \log_v \left( \frac{1}{|\lambda|} \right) \prod_{i=1}^{v} \log_v \left( \frac{1}{|\lambda_i|} \right)} \quad \text{where} \quad \lambda \in \left[ \frac{1}{\exp_v(2)}, \frac{1}{\exp_v(2)} \right] \setminus \{0\}
\]

For any \( v \), using \( P_v^\lambda \) to mix over \( \lambda \) in our proof technique (in conjunction with an appropriate LLN, like Theorem 11) gives the result the optimal iterated-logarithm rate. Furthermore, we show in Section D.2 that as \( v \) increases, the proportionality constant on the result improves. \( P_1^\lambda, P_2^\lambda, \ldots \) require progressively more stringent LLNs (because the support of \( P_v^\lambda \) decreases with increasing \( v \)) but lead to progressively tighter derived LIL bounds.

A similar family of distributions to \( \{P_v^\lambda\}_{v=1,2,\ldots} \) was considered by Robbins and Siegmund ([12], Example 4) in a strictly asymptotic setting. In this section, some arguments made in that paper ([12], Section 4) are extended and generalized to finite times.

D.2 Analysis Sketch with \( P_\lambda^v \)

For completeness, we now more thoroughly discuss the effects of averaging over \( \lambda \) with one of the distributions \( \{P_1^\lambda, P_2^\lambda, P_4^\lambda, \ldots\} \) in place of \( P_1^\lambda \) in the proof of Theorem 7. Note that for any \( v \), \( P_\lambda^v \) has a tractable closed-form antiderivative: for \( \lambda > 0 \), it is \( \frac{1}{\log_v(\frac{1}{\lambda})} \).

Suppose the distribution \( P_\lambda^v \) is used in the proof for some \( v \). The first stage of the proof to prove a uniform LLN bound analogous to Theorem 11 the constants will be different and the initial time condition more restrictive to account for the smaller support of \( P_\lambda^v \) relative to \( P_1^\lambda \), but otherwise this step follows Section 3 closely.

Working within the “good” \((1-\delta)\)-probability event of the resulting LLN, the proof then requires a moment bound analogous to Lemma 3. This is where the averaging distribution \( P_\lambda^v \) plays a role, replacing \( P_1^\lambda \) in the proof of Lemma 3. For any \( v \), Eq. (5) then becomes

\[
\exp \left( \frac{M_2^2}{2U_t} \right) \exp \left( -\frac{1}{2} U_t \left( \frac{M_t}{U_t \sqrt{3}} \right)^2 \right) \times \begin{cases} 
\int \frac{M_t}{U_t} \left( \frac{1}{1+\sqrt{3}} \right) P_\lambda^v(d\lambda) & M_t > 0 \\
\int \frac{M_t}{U_t} \left( \frac{1}{1-\sqrt{3}} \right) P_\lambda^v(d\lambda) & M_t < 0 
\end{cases}
\]

\[
= \exp \left( \frac{M_2^2}{3U_t} \right) \left[ \frac{1}{\log_v \left( \frac{U_t}{|M_t| (1+\sqrt{3})} \right)} - \frac{1}{\log_v \left( \frac{U_t}{|M_t| (1-\sqrt{3})} \right)} \right] P_\lambda^v(d\lambda) 
\]

\[
= \exp \left( \frac{M_2^2}{3U_t} \right) [F_v(\log(S_t)) - F_v(\log(S_t) + \log(\alpha))] \quad (8)
\]

where \( S_t = \frac{U_t}{|M_t| (1+\sqrt{3})} \), \( \alpha = \frac{1+\sqrt{1/3}}{1-\sqrt{1/3}} \) and \( F_v(x) = \frac{1}{\log_{v-1}(x)} \). Note that the derivative of \( F \) is expressible as \( F'_v(x) = -\frac{1}{x \log_{v-1}(x) \prod_{i=1}^{v-1} \log_i(x)} \).
$F_v(\cdot)$ is monotone decreasing and convex, so (8) can be lower-bounded to first order as follows.

$$\text{Eq. (8)} \geq \exp \left( \frac{M^2}{3U_t} \right) \log (\alpha) \left(-F'_v (\log (S_t) + \log (\alpha)) \right)$$

$$= \exp \left( \frac{M^2}{3U_t} \right) \log (\alpha) \frac{1}{\log (\alpha S_t) \log_{v-1} (\log (\alpha S_t)) \left[ \prod_{i=1}^{v-1} \log_i (\log (\alpha S_t)) \right]}$$

$$\geq 2 \exp \left( \frac{M^2}{3U_t} \right) \log_v (\alpha S_t) \left[ \prod_{i=1}^{v} \log_i (\alpha S_t) \right] \quad \text{(9)}$$

Eq. (9) can be compared directly to Eq. (6) in the proof of Lemma 13. Proceeding from (9) and carrying out the rest of the proof of Lemma 13 and Theorem 7, it can be verified that the resulting uniform non-asymptotic LIL bound, for sufficiently high $t$, is at most

$$\sqrt{3U_t \left( \log \left( \frac{2}{\delta} \right) + \sum_{i=2}^{v+1} \log_i (\alpha S_t) + \log_{v+1} (\alpha S_t) \right)}$$

In particular, the result of Theorem 7, with a proportionality constant of $\sqrt{6}$, is recovered for $v = 1$. Also, as $t \to \infty$ the $\log_2 (\alpha S_t)$ term dominates, and it has an unimprovable leading constant $(\sqrt{3})$ for any $v \geq 2$. (An asymptotic version of this was shown in [12].)

E. Proof of Theorem 3

In this section, let $M_t$ be the Rademacher random walk, $k := \frac{1}{\sqrt{2}}$, and $C_1$ be as defined in Theorem 3. Our proof will use submartingales rather than supermartingales; therefore, we will employ a standard optional stopping theorem for submartingales (paralleling Theorem 8):

**Theorem 17** (Optional Stopping for Submartingales (5)). Let $M_t$ be a submartingale. Then if $\tau$ is an a.s. bounded stopping time, $\mathbb{E}[M_{\tau}] \geq \mathbb{E}[M_0]$.

We will also construct a family of exponential submartingales, analogous to the supermartingale construction of Lemma 9 in the upper bound proof.

**Lemma 18.** The process $Z_t^\lambda := \exp (\lambda M_t - k\lambda^2 U_t)$ is a submartingale for $\lambda \in \left[ -\frac{1}{e^2}, \frac{1}{e^2} \right]$.

**Proof.** We rely on the inequality $\cosh(x) \geq e^{kx^2}$ over $x \in \left[ -\frac{1}{e^2}, \frac{1}{e^2} \right]$, so that $\mathbb{E}[Z_t^\lambda | F_{t-1}] = \mathbb{E}[\exp (\lambda \xi_t) | F_{t-1}] e^{-k\lambda^2} Z_{t-1}^\lambda = \cosh(\lambda) e^{-k\lambda^2} Z_{t-1}^\lambda \geq Z_{t-1}^\lambda$. ■

E.1 Preliminaries and Proof Overview

Many aspects of this proof parallel that of the upper bound. Again, the idea is to choose $\lambda$ stochastically from a probability space $(\Omega, \mathcal{F}, P, \lambda)$ such that $P_\lambda (d\lambda) = \frac{d\lambda}{|\lambda| (\log (\frac{1}{|\lambda|}))}$, and the parameter $\lambda$ is chosen independently of the $\xi_1, \xi_2, \ldots$, so that $Z_t^\lambda$ is defined on the product space.

As in the upper bound proof, write $\mathbb{E}^\lambda [\cdot]$ to denote the expectation with respect to $(\Omega, \mathcal{F}, P, \lambda)$. For consistency with previous notation, we continue to write $\mathbb{E} [\cdot]$ to denote the expectation w.r.t. the original probability space $(\Omega, \mathcal{F}, P)$ which encodes the stochasticity of $M_t$.

Just as for the upper bound, our proof of this lower bound reasons about the value of a particular stopping time. However, the stopping time used here is slightly different, because the conditions for convergence of our submartingales are more restrictive than those for supermartingales.
To be concrete, define \( \sigma_\delta := \frac{e^4}{k} \log \left( \frac{2}{\delta} \right) \). For a given finite time horizon \( T > \sigma_\delta \), we use a stopping time defined as

\[
\tau(T) := \min \left\{ t \in [\sigma_\delta, T) : |M_t| > \frac{2kU_t}{e^2} \right\} \lor T
\]

\[
|M_t| \leq \frac{2kU_t}{e^2} \land |M_t| > \sqrt{2kU_t \log \left( \frac{2kU_t}{C_1 \delta} \right) + 2kU_t \log \left( \frac{2kU_t}{|M_t| + 2\sqrt{kU_t}} \right) + 2kU_t \log \left( \frac{2kU_t}{|M_t| + 2\sqrt{kU_t}} \right)}
\]

(10)

We also require a moment bound for the \( \lambda \)-mixed submartingales, whose proof is in Section E.3.

**Lemma 19.** For any \( t \), \( E^\lambda \left[ Z_t^\lambda \right] \leq G_t := \left\{ \begin{array}{ll}
15 \exp \left( \frac{M_T^2}{4U_T} \right) & \text{if } |M_t| \leq \frac{2kU_t}{e^2} \\
7 \exp \left( \frac{M_T^2}{4U_T} \right) & \text{if } |M_t| > \frac{2kU_t}{e^2}
\end{array} \right.
\]

Finally, we use a “one-sided Lipschitz” characterization of \( M_t \) - that \( G_t \) will not grow too fast when \( t \approx \tau(T) \):

**Lemma 20.** For any \( T > \frac{e^4}{k} \log 2 \), \( G_{\tau(T)} \leq \frac{14}{11} G_{\tau(T)-1} \).

With these tools, the proof can be outlined.

**Proof Sketch of Theorem 3.** Here fix \( T \) and write \( \tau = \tau(T) \) as defined in (10). It suffices to prove that \( P(\tau < T) \geq \delta \) under the given assumption on \( \delta \). We have

\[
1 \overset{(a)}{=} E^\lambda \left[ E \left[ Z_{\tau}^\lambda \right] \right] \overset{(b)}{=} E \left[ E^\lambda \left[ Z_{\tau}^\lambda \right] \right] \overset{(c)}{=} E \left[ G_{\tau} \right] \leq E \left[ G_{\tau} \mid \tau < T \right] P(\tau < T) + E \left[ G_{\tau} \mid \tau = T \right] \\
\overset{(d)}{\leq} \frac{14}{11} \left( E \left[ G_{\tau-1} \mid \tau < T \right] P(\tau < T) + E \left[ G_{\tau-1} \mid \tau = T \right] \right)
\]

(11)

where (a) is by Optional Stopping (Theorem 17), (b) is by Tonelli’s Theorem, (c) is by Lemma 19 and (d) is by Lemma 20. The result is then proved by upper-bounding \( E \left[ G_{\tau-1} \mid \tau < T \right] \) and \( E \left[ G_{\tau-1} \mid \tau = T \right] \), using the upper bounds on \( |M_{\tau-1}| \) given by the definition of \( \tau \).

**E.2 Full Proof of Theorem 3.**

**Proof of Theorem 3.** Throughout this proof, fix \( T \) and write \( \tau = \tau(T) \) as defined in (10). It suffices to prove that \( P(\tau < T) \geq \delta \) under the given assumption on \( \delta \).

We do this by working with (11) from the proof sketch; this states that

\[
\frac{11}{14} \leq E \left[ G_{\tau-1} \mid \tau < T \right] P(\tau < T) + E \left[ G_{\tau-1} \mid \tau = T \right]
\]

(12)

By definition of \( \tau \), we have

\[
|M_{\tau-1}| \leq \frac{2k}{e^2} U_{\tau-1} \quad \text{and} \quad |M_{\tau-1}| \leq \sqrt{2kU_{\tau-1} \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}| + 2\sqrt{kU_{\tau-1}}} \right) + 2kU_{\tau-1} \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}| + 2\sqrt{kU_{\tau-1}}} \right)}
\]

(13)
Therefore, from (13),

\[ G_{\tau-1} = \frac{15 \exp \left( \frac{M_{\tau-1}^2}{4kU_{\tau-1}} \right)}{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)} \leq \frac{15}{\sqrt{C_1 \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}} \]  

(14)

So from (14),

\[ E \left[ G_{\tau-1} \mid \tau = T \right] \leq E \left[ \frac{15}{\sqrt{C_1 \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}} \mid \tau = T \right] \leq \frac{15}{\sqrt{C_1}} = \frac{11}{28} \]  

(15)

where the last inequality is by the assumption \( \delta \geq \frac{4}{\log(kU_{\tau-1})} \) and Lemma 21.

Also, from (14),

\[ E \left[ G_{\tau-1} \mid \tau < T \right] \leq E \left[ \frac{15}{\sqrt{C_1 \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}} \mid \tau < T \right] \leq \frac{15}{\sqrt{C_1 \log \left( \frac{2kU_{\tau-1}}{2U_{\tau-1}+2\sqrt{kU_{\tau-1}}} \right)}} \]  

(15)

where (a) uses \( |M_{\tau-1}| \leq \frac{2k}{e^2} U_{\tau-1} \) (by (13)) and (b) uses \( U_{\tau-1} = \tau - 1 \geq \frac{e^4}{k} \log 2 - 1 \).

Substituting this and (15) into (12) gives \( 1 \leq \left( \frac{15}{28} \right) P(\tau < T) + \frac{1}{2} \). Therefore, \( P(\tau < T) \geq \delta \), finishing the proof.

The proof of Theorem 3 requires a supporting lemma, which is proved in Section E.4.

Lemma 21. Within the event \{\( \tau(T) = T \)\}, if \( \delta \geq \frac{4}{\log(kU_{\tau-1})} \), then

\[
\frac{1}{\sqrt{\delta \log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}} \leq 1
\]

E.3 Proof of Lemma 19

Proof of Lemma 19. First note that

\[
\mathbb{E}^{\lambda} \left[ Z_{\lambda}^t \right] = \int_{-1/e^2}^0 \exp \left( \lambda M_t - k\lambda^2 U_t \right) \frac{d\lambda}{-\lambda \left( \log \frac{1}{\lambda} \right)^2} + \int_{0}^{1/e^2} \exp \left( \lambda M_t - k\lambda^2 U_t \right) \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2}
\]
\begin{align*}
\mathbb{E}^\lambda \left[ Z_t^\lambda \right] &\leq 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
&\leq 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left[ \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \right] \\
&= 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left[ \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \right]
\end{align*}

Suppose that \( \frac{|M_t|}{2kU_t} \leq \frac{1}{e^2} \). Define an integer \( N \) such that
\begin{equation}
\frac{|M_t|}{2kU_t} + \sqrt{\frac{N - 1}{kU_t}} \leq \frac{1}{e^2} \leq \frac{|M_t|}{2kU_t} + \sqrt{\frac{N}{kU_t}}
\end{equation}
(Note that we are guaranteed \( N \geq 1 \) by the assumption \( \frac{|M_t|}{2kU_t} \leq \frac{1}{e^2} \)). Combining this with (16),
\begin{align*}
\mathbb{E}^\lambda \left[ Z_t^\lambda \right] &\leq 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left[ \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \sum_{i=0}^{N-1} \int_{\frac{|M_t|}{2kU_t} + \sqrt{\frac{i}{kU_t}}}^{\frac{|M_t|}{2kU_t} + \sqrt{\frac{i+1}{kU_t}}} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \right] \\
&= 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left[ \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} + \sum_{i=0}^{N-1} \int_{\frac{|M_t|}{2kU_t} + \sqrt{\frac{i}{kU_t}}}^{\frac{|M_t|}{2kU_t} + \sqrt{\frac{i+1}{kU_t}}} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \right]
\end{align*}
Now we have
\begin{align*}
\int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \leq \int_0^{1/e^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} = \frac{1}{\log \frac{1}{\lambda}} \bigg| \frac{1}{\log \frac{1}{\lambda}} \bigg| = \frac{1}{\log \frac{2kU_t}{|M_t|}}
\end{align*}
Substituting this and Lemma 22 into (18), we get
\begin{align*}
\mathbb{E}^\lambda \left[ Z_t^\lambda \right] &= 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left[ \frac{1}{\log \frac{2kU_t}{|M_t|}} + \frac{1}{\log \left( \frac{2kU_t}{|M_t|+2\sqrt{kU_t}} \right)} \left( \frac{e}{e - 1} + \frac{3}{2 \log \left( \frac{e^2}{1+e^2/k} \right)} \right) \right] \\
&\leq \frac{e^2 \exp \left( \frac{M_t^2}{4kU_t} \right)}{\log \left( \frac{2kU_t}{|M_t|+2\sqrt{kU_t}} \right)} \left( \frac{6}{\log \left( \frac{e^2}{1+e^2/k} \right)} \right) \leq \frac{15 \exp \left( \frac{M_t^2}{4kU_t} \right)}{\log \left( \frac{2kU_t}{|M_t|+2\sqrt{kU_t}} \right)} = G_t
\end{align*}
which proves the result when \( |M_t| \leq \frac{2kU_t}{e^2} \).

Alternatively, if \( |M_t| > \frac{2kU_t}{e^2} \), from (16) we have
\begin{align*}
\mathbb{E}^\lambda \left[ Z_t^\lambda \right] &\leq 2 \exp \left( \frac{M_t^2}{4kU_t} \right) \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \\
&\leq 2e^2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left( \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} d\lambda \right) \left( \int_0^{1/e^2} \frac{d\lambda}{\lambda \left( \log \frac{1}{\lambda} \right)^2} \right) \\
&= e^2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left( \int_0^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} d\lambda \right) \leq e^2 \exp \left( \frac{M_t^2}{4kU_t} \right) \int_{-\infty}^{1/e^2} e^{-kU_t \left( \lambda - \frac{|M_t|}{2kU_t} \right)^2} d\lambda \\
&= e^2 \exp \left( \frac{M_t^2}{4kU_t} \right) \left( \sqrt{\frac{\pi}{2\sqrt{kU_t}}} \right) \leq e^2 \exp \left( \frac{M_t^2}{4kU_t} \right) \frac{\sqrt{\pi}}{2 \log \left( \sqrt{kU_t} \right)} \leq \frac{7 \exp \left( \frac{M_t^2}{4kU_t} \right)}{\log \left( \frac{2kU_t}{\sqrt{kU_t}} \right)}
\end{align*}
where (a) is by Chebyshev’s integral inequality (Lemma 23), and (b) is because \( \frac{\log \left( \sqrt{kU_t} \right)}{\sqrt{kU_t}} \leq 1 \).
Lemma 22. Define $N$ as in (17). Then

$$
\sum_{i=0}^{N-1} \int_{\frac{|M_t|}{2kU_t} + \frac{\sqrt{N}}{kU_t}} \exp\left(-kU_t \left(\lambda - \frac{|M_t|}{2kU_t}\right)^2\right) \frac{d\lambda}{\lambda \left(\log \frac{1}{\lambda}\right)^2} \leq \frac{1}{\log \left(\frac{1}{\mu + \sigma}\right)} \left(\frac{e}{e - 1} + \frac{3}{2 \log \left(\frac{e}{1+e^2/\sqrt{e}}\right)}\right)
$$

Proof. For convenience, define $\mu := \frac{|M_t|}{2kU_t}$ and $\sigma := \frac{1}{\sqrt{kU_t}}$. In particular, this means that $\mu + \sigma \sqrt{N - 1} \geq \frac{1}{e} \leq \mu + \sigma \sqrt{N}.$

$$
\sum_{i=0}^{N-1} \int_{\frac{|M_t|}{2kU_t} + \frac{\sqrt{N}}{kU_t}} \exp\left(-kU_t \left(\lambda - \frac{|M_t|}{2kU_t}\right)^2\right) \frac{d\lambda}{\lambda \left(\log \frac{1}{\lambda}\right)^2} \leq \sum_{i=0}^{N-1} \int_{\mu + \sigma \sqrt{i+1}} \frac{d\lambda}{\lambda \left(\log \frac{1}{\lambda}\right)^2} = \sum_{i=0}^{N-1} \exp\left(-\frac{(\sigma \sqrt{i+1})^2}{\mu + \sigma \sqrt{i+1}}\right) \frac{d\lambda}{\lambda \left(\log \frac{1}{\lambda}\right)^2} \leq \frac{1}{\log \left(\frac{1}{\mu + \sigma}\right)} \sum_{i=0}^{N-1} e^{-i} \left(1 + \log \left(\frac{\mu + \sigma \sqrt{i+1}}{\mu + \sigma}\right)\right) \leq \frac{1}{\log \left(\frac{1}{\mu + \sigma}\right)} \left(1 + \frac{1}{\log \left(\frac{e}{1+e^2/\sqrt{e}}\right)}\right) \sum_{i=0}^{\infty} e^{-i} \log \left(1 + \sqrt{i}\right) \leq \frac{1}{\log \left(\frac{1}{\mu + \sigma}\right)} \left(1 + \frac{3}{2 \log \left(\frac{e}{1+e^2/\sqrt{e}}\right)}\right)
$$

where (19) follows from Lemma 23.

Lemma 23. $\sum_{i=0}^{\infty} e^{-i} \log \left(1 + \sqrt{i}\right) \leq \frac{3}{2}$

Proof. Take $f(x) = e^{-x/2} \log (1 + \sqrt{x})$ for $x \geq 0$. Then $f'(x) = \frac{e^{-x/2}}{2} \left(\frac{1}{\sqrt{x(1+\sqrt{x})}} - \log (1 + \sqrt{x})\right).$ Since $\frac{1}{\sqrt{x(1+\sqrt{x})}}$ is monotone decreasing and $\log (1 + \sqrt{x})$ is monotone increasing, $f'(x)$ has exactly one root, corresponding to the maximum of $f(x)$. This can be numerically confirmed to occur at $x^* \approx 0.745$, and $f(x^*) \leq 0.5$.

So $\sum_{i=0}^{\infty} e^{-i/2} \log (1 + \sqrt{i}) \leq \sum_{i=0}^{\infty} e^{-i/2} \log (1 + \sqrt{i}) \leq \frac{1}{2} \sum_{i=0}^{\infty} e^{-i/2} = \frac{\sqrt{e}}{2(e-1)} \leq \frac{3}{2}.$
E.4 Ancillary Results and Proofs

*Proof of Lemma 20* Note that by definition of $\tau$,

$$|M_{\tau-1}| \leq \frac{2k}{e^2} U_{\tau-1} \quad \text{and} \quad |M_{\tau-1}| \leq \sqrt{2kU_{\tau-1} \log \left( \frac{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}{C_1 \delta} \right)} \quad (20)$$

Firstly, we have

$$\frac{\exp \left( \frac{M_U^2}{4kU_{\tau-1}} \right)}{\exp \left( \frac{M_U^2}{4kU_{\tau-1}} \right)} \leq \frac{\exp \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}{\exp \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)} \leq \frac{\exp \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau}} \log |M_{\tau-1}|+2\sqrt{kU_{\tau}} \log (13/12) \right)}{\exp \left( \frac{2kU_{\tau}}{|M_{\tau-1}|+2\sqrt{kU_{\tau}}} \right)}$$

where (a) is because $M_{\tau-1} \xi \leq |M_{\tau-1}| \leq \frac{2kU_{\tau-1}}{e^2}$ by [20], and because $\frac{2kU_{\tau-1}}{4kU_{\tau-1}} = \frac{1}{e^2} \leq \frac{1}{4k(108)} = \frac{1}{11}.$

We write $\tau(T)$ more concisely as $\tau$ here, and note that the minimum value of $T$ implies that $\tau(T) \geq 108$, a fact we will use throughout the proof. Also, we have

$$\log \left( \frac{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}{\log \left( \frac{2kU_{\tau}}{|M_{\tau-1}|+2\sqrt{kU_{\tau}}} \right)} \right) \leq \frac{\log(3)}{11} \leq 12$$

where (a) is because $f(x) = \log \left( \frac{x}{C+\sqrt{2x}} \right)$ is monotone increasing for any $C \geq 0$; and (b) is because $|M_{\tau-1}|+2\sqrt{kU_{\tau}} \geq 2\sqrt{kU_{\tau}} \geq 12$, so $1+|M_{\tau-1}|+2\sqrt{kU_{\tau}} \leq \frac{13}{12} (|M_{\tau-1}|+2\sqrt{kU_{\tau}})$.

Now we have

$$\frac{2kU_{\tau}}{|M_{\tau-1}|+2\sqrt{kU_{\tau}}} \leq \frac{2kU_{\tau}}{2kU_{\tau-1}+2\sqrt{kU_{\tau}}} \geq \frac{kU_{\tau}}{2kU_{\tau-1}+2\sqrt{kU_{\tau}}} \geq \min \left( \frac{e^2}{2k}, \frac{1}{2} \sqrt{kU_{\tau}} \right) \geq 3$$

where (b) uses [20] and (c) uses $U_{\tau} = \tau \geq 108$. This means that from (22),

$$\frac{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}{\log \left( \frac{2kU_{\tau}}{|M_{\tau-1}|+2\sqrt{kU_{\tau}}} \right)} \leq \frac{\log(3)}{11} \leq 12$$

Note that by [20], $|M_{\tau-1}| \leq \frac{2k}{e^2} U_{\tau-1}$, so $G_{\tau-1} = \frac{15 \exp \left( \frac{M_U^2}{4kU_{\tau-1}} \right)}{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}$.

Suppose $|M_{\tau}| \leq \frac{2k}{e^2} U_{\tau}$. Then using Lemma 19 [21], and [23],

$$\frac{G_{\tau}}{G_{\tau-1}} = \frac{\exp \left( \frac{M_U^2}{4kU_{\tau}} \right)}{\exp \left( \frac{M_U^2}{4kU_{\tau-1}} \right)} \leq \frac{\log \left( \frac{2kU_{\tau-1}}{|M_{\tau-1}|+2\sqrt{kU_{\tau-1}}} \right)}{\log \left( \frac{2kU_{\tau}}{|M_{\tau}|+2\sqrt{kU_{\tau}}} \right)} \leq \frac{7}{6} \frac{12}{11} = \frac{14}{11}$$
Alternatively, if $|M_{\tau}| > \frac{2k}{e}U_{\tau}$, we can use Lemma 19 and (21) to conclude that

$$
G_{\tau} = \frac{7 \exp\left(\frac{M_{\tau}^2}{4kU_{\tau}}\right)}{15 \exp\left(\frac{M_{\tau}^2}{4kU_{\tau}}\right)} \log\left(\frac{2kU_{\tau}}{|M_{\tau}| + 2\sqrt{kU_{\tau}} - 1}\right) \leq \frac{\exp\left(\frac{M_{\tau}^2}{4kU_{\tau}}\right)}{\exp\left(\frac{M_{\tau}^2}{4kU_{\tau}}\right)} \log\left(\frac{\sqrt{kU_{\tau}}}{kU_{\tau}}\right) \leq \frac{7}{6}(1) \leq \frac{14}{11}.
$$

Proof of Lemma 21 The definition of $\tau(= T)$ and the fact that $|M_{\tau-1}| \geq 0$ imply that

$$
|M_{\tau-1}| = |M_{\tau-1}| \leq \sqrt{2kU_{\tau-1} \log\left(\frac{2kU_{\tau-1}}{|M_{\tau-1}| + 2\sqrt{kU_{\tau-1}} - 1}\right)} \leq \sqrt{2kU_{\tau-1} \log\left(\frac{\log(kU_{\tau-1})}{2C_1}\right)}
$$

Consequently,

$$
\log\left(\frac{2kU_{\tau-1}}{|M_{\tau-1}| + 2\sqrt{kU_{\tau-1}} - 1}\right) \geq \log\left(\frac{2kU_{\tau-1}}{\sqrt{2kU_{\tau-1} \log\left(\frac{\log(kU_{\tau-1})}{2C_1}\right) + 2\sqrt{kU_{\tau-1}}}}\right)
$$

$$
= \log\left(\frac{\sqrt{2kU_{\tau-1}}}{\sqrt{\log\left(\frac{\log(kU_{\tau-1})}{2C_1}\right) + 2}}\right) \geq \log\left(\frac{\sqrt{2kU_{\tau-1}}}{\sqrt{\log\left(\log^2(kU_{\tau-1})\right) + 2}}\right)
$$

$$
\geq \log\left(\frac{\sqrt{2kU_{\tau-1}}}{(kU_{\tau-1})^{1/16} + 1}\right) \geq \frac{1}{4} \log(kU_{\tau-1})
$$

where (a) uses that $\delta \geq \frac{4}{\log(kU_{\tau-1})}$ and $8C_1 \geq 1$, (b) uses that $\log(kU_{\tau-1}) \leq (kU_{\tau-1})^{1/8}$, and (c) uses that $(kU_{\tau-1})^{1/16} + 1 \leq (kU_{\tau-1})^{1/4}$ for $kU_{\tau-1} = k(\tau - 1) \geq e^4$ for $k \geq 26$. Therefore,

$$
\frac{1}{\sqrt{\delta \log\left(\frac{2kU_{\tau-1}}{|M_{\tau-1}| + 2\sqrt{kU_{\tau-1}} - 1}\right)}} \leq \frac{1}{\sqrt{\frac{4}{15} \log(kU_{\tau-1})}} \leq 1
$$

after substituting (21) and again using $\delta \geq \frac{4}{\log(kU_{\tau-1})}$.

Lemma 24 (Chebyshev’s Integral Inequality). If $f(x)$ and $g(x)$ are respectively monotonically increasing and decreasing functions over an interval $(a, b)$, and $\int_a^b f(x)dx$ and $\int_a^b g(x)dx$ are both defined and finite, then

$$
\int_a^b f(x)g(x)dx \leq \frac{1}{b - a} \left(\int_a^b f(x)dx\right) \left(\int_a^b g(x)dx\right)
$$

Proof. By the monotonicity properties of the functions, we know for any $x, y \in (a, b)$ that $(f(x) - f(y))(g(x) - g(y)) \leq 0$. Therefore,

$$
0 \geq \int_a^b \int_a^b (f(x) - f(y))(g(x) - g(y))dydx = 2(b - a) \int_a^b f(x)g(x)dx - 2 \left(\int_a^b f(x)dx\right) \left(\int_a^b g(x)dx\right)
$$

which yields the result upon simplification.