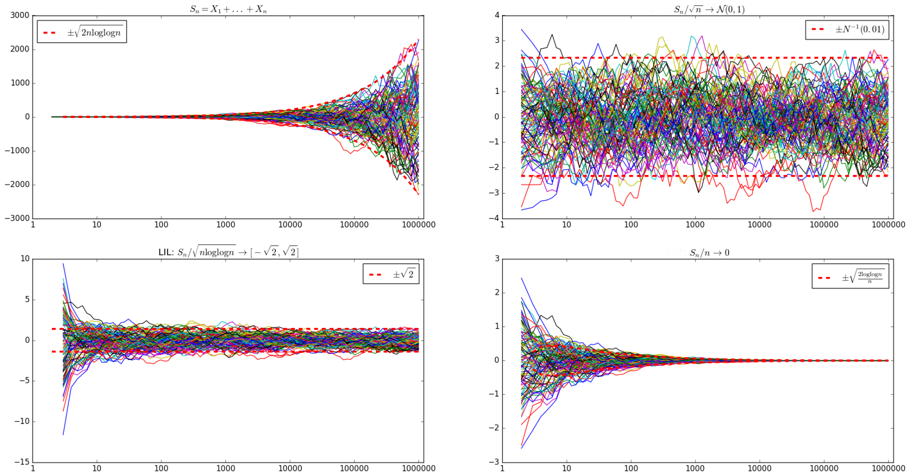

Large Deviations and the Law of the Iterated Logarithm

Hanan Ather[*,†]

MAT 5171

Probability Theory II

April 3, 2024

University of Ottawa

[*]Statistics Canada, Modern Statistical Methods and Data Science Division
[†]University of Ottawa, Department of Mathematics and Statistics

Contents

§1 Moment Generating Functions

Definition 1.1 (Moment generating function) Let X be a discrete random variable that can take on a finite number l of distinct values $x_1, x_2, \ldots, x_l$, each with a corresponding probability $p_1, p_2, \ldots, p_l$. These probabilities must satisfy the conditions $0 \leq p_i \leq 1$ and $\sum_{i=1}^{l} p_i = 1$. The **moment generating function (MGF)** of X, denoted $M_X(t)$, is a function defined for all real t as the expected value of the exponential function e^{tX}, and is given by the formula:

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{i=1}^{l} p_i e^{tx_i}.$$

The MGF can be considered a transformation of the probability distribution of X, reflecting the distribution's behavior through the moments it generates. For example, the n-th moment of X, μ'_n, is obtained by differentiating $M_X(t)$ n times with respect to t and evaluating at $t = 0$:

$$\mu'_n = \left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0}.$$

If two random variables have the same MGF, and it exists in an open interval around $t = 0$, they have the same distribution. The MGF is not guaranteed to exist for all random variables, but when it does, it offers a powerful tool for analysis. Additionally, the MGF is closely related to other transforms in probability theory, such as the characteristic function.

Assuming $c = \max |x_i|$, the series expansion of e^{tX} converges and is bounded by $e^{|t|c}$, thus ensuring the existence of the moment generating function $M(t)$ for the random variable X. The moment generating function can be expressed as an infinite Taylor series:

$$M(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k]. \tag{1}$$

This expansion allows us to write the k-th moment of X as the k-th derivative of $M(t)$ evaluated at zero:

$$\mathbb{E}[X^k] = M^{(k)}(0). \tag{2}$$

Moreover, taking the k-th derivative of the MGF with respect to t and evaluating it at $t = 0$ provides us with a direct calculation of the moments of X, which is one reason why $M(t)$ is a valuable tool in probability theory. For the simple random variable X, this derivative is:

$$M^{(k)}(t) = \sum_{i=1}^{l} p_i x_i^k e^{tx_i} = \mathbb{E}[X^k e^{tX}]. \tag{3}$$

It is important to note that $M(0) = 1$, as the zeroth moment of a random variable is always 1, corresponding to the total probability.

Example 1.2. Consider a Bernoulli random variable X which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$. The moment generating function (MGF) of X, denoted by $M_X(t)$, is defined as the expected value of e^{tX}, and is given by:

$$M_X(t) = \mathbb{E}[e^{tX}] = p \cdot e^{t \cdot 1} + q \cdot e^{t \cdot 0} = pe^t + q.$$

The first moment, or the expected value, is the first derivative of the MGF evaluated at $t = 0$:

$$M_X'(t) = \frac{d}{dt}(pe^t + q) = pe^t, \quad \text{so} \quad M_X'(0) = pe^0 = p.$$

The second moment is obtained by differentiating again to obtain the second derivative, and evaluating at $t = 0$:

$$M_X''(t) = \frac{d^2}{dt^2}(pe^t + q) = pe^t, \quad \text{so} \quad M_X''(0) = pe^0 = p.$$

The variance of X is then calculated as:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = M_X''(0) - (M_X'(0))^2 = p - p^2 = pq.$$

Consider n independent random variables $X_1, X_2, \ldots, X_n$. Independence implies that for any real number t, the exponential functions $e^{tX_1}, e^{tX_2}, \ldots, e^{tX_n}$ are also independent. Let $M(t)$ represent the moment generating function of the sum $S = X_1 + X_2 + \ldots + X_n$, and let $M_{X_i}(t)$ denote the moment generating function of each X_i.

Since moment generating functions are defined as the expected value of the exponential of the random variable, for independent random variables, the expected value of a product is the product of the expected values, due to their independence. This leads to the following relation for the MGF of the sum S:

$$M_S(t) = \mathbb{E}[e^{tS}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] = \prod_{i=1}^{n} M_{X_i}(t), \tag{4}$$

where $M_S(t)$ is the MGF of S and $M_{X_i}(t)$ is the MGF of X_i[1].

Definition 1.3 (Cumulant Generating Function) The *cumulant generating function* of X (or of its distribution) is

$$C(t) = \log M(t) = \log \mathbb{E}[e^{tX}]^{a}. \tag{5}$$

[a]Note that $M(t)$ is strictly positive.

Since $C' = \frac{M'}{M}$ and $C'' = \left(\frac{M'M''}{M^2}\right) - \left(\frac{M'}{M}\right)^2$, and since $M(0) = 1$,

$$C(0) = 0, \quad C'(0) = \mathbb{E}[X], \quad C''(0) = \text{Var}[X]. \tag{6}$$

Let $m_k = \mathbb{E}[X^k]$. The leading term in (9.2) is $m_0 = 1$, and so a formal expansion of the logarithm in (5) gives

$$C(t) = \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu+1}}{\nu} \left(\sum_{k=1}^{\infty} \frac{m_k}{k!} t^k\right)^{\nu}. \tag{7}$$

Since $M(t) \to 1$ as $t \to 0$, this expression is valid for t in some neighborhood of 0. By the theory of series, the powers on the right can be expanded and terms with a common factor t^i are grouped together in the expansion of the cumulant generating function of X as

$$C(t) = \sum_{i=1}^{\infty} \frac{c_i}{i!} t^i, \tag{8}$$

which is convergent within a certain neighborhood of 0.

Here, c_i represents the cumulants of X. By matching terms in the expansions of the cumulant and moment generating functions and utilizing the previously established relationships, we deduce that $c_1 = m_1$ and $c_2 = m_2 - m_1^2$. Although expressing each c_i in terms of the moments

[1]This property allows for the simplification of calculating the MGF of the sum of independent random variables by merely multiplying their individual MGFs, which is a convenient method especially when dealing with sums of many random variables. Furthermore, distributions can, in principle, be recovered from their MGFs, provided that the MGF uniquely identifies the distribution, which is the case if the MGF exists in an interval around $t = 0$.

$m_1, \ldots, m_i$ is possible, it can quickly become complex. When $\mathbb{E}[X] = 0$, we find $m_1 = c_1 = 0$, and straightforward computation shows that

$$c_3 = m_3, \quad c_4 = m_4 - 3m_2^2. \tag{9}$$

Logarithms transform the product relation into a sum,

$$C(t) = C_1(t) + \ldots + C_n(t), \tag{10}$$

corresponding to the sum of the cumulant generating functions, which is applicable when independence is assumed. The addition property of cumulants for independent random variables follows from this and definition (6).

It is evident that $M''(t) = \mathbb{E}[X^2 e^{tX}] \geq 0$, and since $(M'(t))^2 \leq M(t)M''(t)$ by the Cauchy-Schwarz inequality, it holds that $C''(t) \geq 0$.

Therefore, both the moment generating function and the cumulant generating function are convex.

§2 Large Deviations

Our goal will be prove the following theorem:

Theorem 2.1 — Suppose that Y satisfies $E[Y] < 0$ and $\mathbb{P}(Y > 0) > 0$. Define ρ and τ by

$$\inf_t M(t) = M(\tau) = \rho, \quad 0 < \rho < 1, \tau > 0,$$

Let Z be a random variable with distribution

$$\mathbb{P}(Z \geq y_j) = e^{\tau y_j} \frac{\mathbb{P}(Y = y_j)}{\rho},$$

and define $E[Z]$, $S^2 = E[Z^2]$ as

$$E[Z] = \frac{M'(\tau)}{\rho} = 0, \quad S^2 = E[Z^2] = \frac{M''(\tau)}{\rho} > 0.$$

Then

$$\mathbb{P}(Y \geq 0) = \rho e^{-\theta},$$

where θ satisfies

$$0 \leq \theta \leq \frac{\tau S}{\mathbb{P}(Z \geq 0)} - \log \mathbb{P}(Z \geq 0).$$

Proof. Consider a simple random variable Y that takes on values y_j with corresponding probabilities p_j. To estimate $P(Y \geq \alpha)$ when Y has a mean of 0 and α is positive, we subtract α from Y to instead estimate $P(Y \geq 0)$ where Y has a negative mean. We assume $E[Y] < 0$ and $P(Y > 0) > 0$, to circumvent trivial cases. The moment generating function $M(t) = \sum_j p_j e^{ty_j}$ satisfies $M'(0) < 0$.

As t approaches infinity, $M(t)$ increases without bound, and since $M(t)$ is convex, it achieves its minimum ρ at some positive τ:

$$\inf_t M(t) = M(\tau) = \rho, \quad 0 < \rho < 1, \quad \tau > 0.$$

An auxiliary random variable Z is defined on a different probability space such that

$$P[Z = y_j] = \frac{e^{\tau y_j}}{\rho} P[Y = y_j],$$

for each y_j in the range of Y, ensuring that the probabilities sum to one. The moment generating function for Z is

$$E[e^{tZ}] = \sum_j \frac{e^{\tau y_j}}{\rho} p_j e^{t y_j} = \frac{M(\tau + t)}{\rho},$$

and hence,

$$E[Z] = \frac{M'(\tau)}{\rho} = 0, \quad s^2 = E[Z^2] = \frac{M''(\tau)}{\rho} > 0.$$

Using Markov's inequality, for all positive t, $P(Y \geq 0)$ can be bounded by ρ. Obtaining inequalities in the opposite direction is more challenging. If Σ' represents the summation over indices j for which $y_j \geq 0$, we have

$$P[Y \geq 0] = \sum{}' p_j = \sum{}' e^{-\tau y_j} p_j e^{\tau y_j} P[Z = y_j].$$

Upon simplifying the sum as $e^{-\theta}$ with $\rho = P[Z \geq 0]$ and invoking Jensen's inequality,

$$-\theta \geq \log \sum{}' e^{-\tau y_j} p_j^{-1} P[Z = y_j] + \log \rho \geq -\tau s^{-1} \sum{}' \frac{y_j}{s} P[Z = y_j] + \log \rho.$$

Lyapounov's Inequality states that for a non-negative random variable X and $0 < \alpha < \beta$, the following holds:

$$(E[|X|^\alpha])^{1/\alpha} \leq (E[|X|^\beta])^{1/\beta}$$

In the given context, Lyapounov's inequality is used to relate the first and second moments of the random variable Z. By setting $\alpha = 1$ and $\beta = 2$, we apply the inequality as follows:

$$E[|Z|] \leq (E[Z^2])^{1/2}$$

Given that $E[Z^2] = s^2$ and $E[Z] = 0$, this simplifies to:

$$E[|Z|] \leq s$$

Applying Lyapounov's inequality leads to a relationship between θ and the bounds for $P[Z \geq 0]$:

$$0 \leq \theta \leq \frac{\tau s}{P[Z \geq 0]} - \log P[Z \geq 0]. \tag{11}$$

□

To use (11) requires a lower bound for $\mathbb{P}[Z \geq 0]$.

Theorem 2.2 — If $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = s^2$, and $\mathbb{E}[Z^4] = \xi^4 > 0$, then $P[Z \geq 0] \geq \frac{s^4}{4\xi^4}$.

Proof. Define Z^+ as Z conditioned on $Z \geq 0$ and Z^- as $-Z$ conditioned on $Z < 0$. Both Z^+ and Z^- are nonnegative, and we let Z be their difference, $Z = Z^+ - Z^-$, and Z^2 be the sum of their squares, $Z^2 = (Z^+)^2 + (Z^-)^2$, which gives us

$$s^2 = \mathbb{E}\left[(Z^+)^2\right] + \mathbb{E}\left[(Z^-)^2\right].$$

Let $p = \mathbb{P}[X \geq 0]$ and using the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[(Z^+)^2\right] = \mathbb{E}\left[\mathbb{1}_{\{Z\geq 0\}} Z^2\right] \leq \mathbb{E}^{1/2}\left[\mathbb{1}^2_{\{Z\geq 0\}}\right] \mathbb{E}^{1/2}\left[Z^4\right] = p^{1/2} s^2.$$

Applying Hölder's inequality, for Z^-, we have

$$\mathbb{E}\left[(Z^-)^2\right] = \mathbb{E}\left[(Z^-)^{3/2}(Z^-)^{1/2}\right] \leq \mathbb{E}^{2/3}\left[(Z^-)^3\right] \mathbb{E}^{1/3}\left[(Z^-)^2\right] \leq s^{2/3}\mathbb{E}^{4/3}\left[Z^-\right].$$

Since the expectation of Z is zero, applying Hölder's inequality again, we find

$$\mathbb{E}\left[(Z^-)^2\right] = \mathbb{E}\left[(Z^+)^2\right] = \mathbb{E}\left[Z^2 \mathbb{1}_{\{Z\geq 0\}}\right] \leq \mathbb{E}^{1/4}\left[Z^4\right]\mathbb{E}^{3/4}\left[\mathbb{1}^4_{\{Z\geq 0\}}\right] = s^{3/4}p^{3/4}.$$

Combining these inequalities gives

$$s^2 \geq p^{1/2}s^2 + s^{2/3}p^{4/3},$$

which concludes the proof. □

§3 The Law of the Iterated Logarithm

Theorem 3.1 — Let $S_n = X_1 + \cdots + X_n$, where the X_n are independent and identically distributed simple random variables with mean 0 and variance 1. If a_n are constants satisfying

$$a_n \to \infty, \quad \text{and} \quad \frac{a_n}{\sqrt{n}} \to 0, \tag{12}$$

then

$$P\left[S_n \geq a_n\sqrt{n}\right] = e^{-a_n^2(1+\xi_n)/2} \tag{13}$$

for a sequence ξ_n going to 0.

⋆ **Intuition:**

$$P\left[S_n \geq a_n\sqrt{n}\right] = e^{-\frac{a_n^2}{2}(1+\xi_n)}$$

indicates that the probability of S_n being greater than $a_n\sqrt{n}$ decays exponentially as n grows large. The term $e^{-\frac{a_n^2}{2}(1+\xi_n)}$ is an exponential function of $-a_n^2$, which itself grows without bound since $a_n \to \infty$. The decay is also affected by $(1+\xi_n)$, where ξ_n is a sequence that goes to 0, slightly adjusting the rate of decay, but not changing the fact that it is exponential.

Now we prove the statement:

Proof. Put $Y_n = S_n - a_n/\sqrt{n} = \sum_{k=1}^n (X_k - a_n/\sqrt{n})$. Then $\mathbb{E}[Y_n] < 0$. X_1 has mean 0 and variance 1, $\mathbb{P}[X_1 > 0] > 0$, and it follows by assumption that $\mathbb{P}[X_1 > a_n/\sqrt{n}] > 0$ for n sufficiently large, in which case $\mathbb{P}[Y_n > 0] \geq \mathbb{P}[X_1 - a_n/\sqrt{n} > 0]$. Thus Theorem 2.1 applies to Y_n for all large enough n.

Let $M_n(t), \rho_n, \tau_n$, and Z_n be associated with Y_n as in the theorem. If $m(t)$ and $c(t)$ are the moment and cumulant generating functions of the X_n, then $M_n(t)$ is the nth power of the moment generating function $e^{-ta_n/\sqrt{n}}m(t)$ of $X_1 - a_n/\sqrt{n}$, and so Y_n has cumulant generating function

$$C_n(t) = -ta_n\sqrt{n} + nc(t).$$

Since τ_n is the unique minimum of $C_n(t)$, and since $C_n''(t) = -a_n\sqrt{n} + nc''(t)$, τ_n is determined by the equation $c'(\tau_n) = a_n/\sqrt{n}$. Since X_1 has mean 0 and variance 1, it follows that $c(0) = c'(0) = 0$, $c''(0) = 1$.

Now $c'(t)$ is nondecreasing because $c(t)$ is convex, and since $c'(\tau_n) = a_n/\sqrt{n}$ goes to 0, τ_n must therefore go to 0 as well and must in fact be $\mathcal{O}(a_n/\sqrt{n})$. By the second-order mean-value theorem for $c'(t)$, $a_n/\sqrt{n} = c'(\tau_n) = \tau_n + \mathcal{O}(\tau_n^2)$, from which follows

$$\tau_n = \frac{a_n}{\sqrt{n}} + \mathcal{O}\left(\frac{a_n^2}{n}\right).$$

By the third-order mean-value theorem for $c(t)$, we have

$$\begin{aligned}\log p_n = C_n(\tau_n) &= -\tau_n a_n\sqrt{n} + nc(\tau_n)\\ &= -\tau_n a_n\sqrt{n} + n\left[\frac{1}{2}\tau_n^2 + O(\tau_n^3)\right].\end{aligned}$$

Applying the result for τ_n gives

$$\log p_n = -\frac{1}{2}a_n^2 + o(a_n^2).$$

Now, considering the moment generating function $M_n(\tau_n, t)/\rho_n$ and the cumulant generating function $D_n(t) = C_n(\tau_n + t) - \log p_n$, we find that the mean of Z_n is $D_n'(0) = 0$. Its variance s_n^2 is $D_n''(0)$, which is

$$s_n^2 = nc''(\tau_n) = n[c''(0) + O(\tau_n)] = n(1 + o(1)).$$

The fourth cumulant of Z_n is $D_n^{(4)}(0) = nc^{(4)}(\tau_n) = O(n)$. Thus, for all sufficiently large n, there is a positive lower bound for the probability that $Z_n \geq 0$. By Theorem 9.1, $P[Y \geq 0] = p_n e^{-\theta_n}$, with $\theta_n = O(a_n) = o(a_n^2)$, and it follows that $P[Y \geq 0] = e^{-a_n^2(1+o(1))/2}$.

□

Theorem 3.2 (Law of the Iterated Logarithm) — Let $S_n = X_1 + \cdots + X_n$, where the X_n are independent, identically distributed simple random variables with mean 0 and variance 1. Then

$$P\left[\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right] = 1. \tag{14}$$

Equivalent to (14) is the assertion that for positive ε

$$P\left[S_n \geq (1+\varepsilon)\sqrt{2n\log\log n}\; i.o.\right] = 0 \tag{15}$$

and

$$P\left[S_n \geq (1-\varepsilon)\sqrt{2n\log\log n}\; i.o.\right] = 1. \tag{16}$$

The set in (14) is, in fact, the intersection over positive rational ε of the sets in (16) minus the union over positive rational ε of the sets in (15).

⋆ Theorem 3.2 states that for a sequence of i.i.d. random variables $\{X_i\}$ with mean 0 and variance 1, the limit superior of the normalized sum $\frac{S_n}{\sqrt{2n\log\log n}}$ as n tends to infinity is 1 with probability 1. Formally,

$$P\left[\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right] = 1.$$

This means that as you consider larger and larger sums of these random variables, the sequence of sums, when scaled by $\sqrt{2n\log\log n}$, will not indefinitely exceed 1, and it will be arbitrarily close to 1 infinitely often.

Proof. We prove the proof in three parts:

$$\phi(n) = \sqrt{2n\log\log n}. \tag{17}$$

If $A_n^{\pm} = \{S_n \geq (1 \pm \varepsilon)\phi(n)\}$, then by (3.1), $P(A_n^{\pm})$ is near $(\log n)^{-(1\pm\varepsilon)^2}$. The theorem (3.1) indicates that the probability of S_n being greater than $a_n\sqrt{n}$ can be approximated by $e^{-\frac{a_n^2}{2}(1+\xi_n)}$, with ξ_n going to 0. If we take a_n to be $(1\pm\varepsilon)\sqrt{2\log\log n}$, then $a_n\sqrt{n}$ becomes $(1\pm\varepsilon)\sqrt{2n\log\log n}$, or $(1\pm\varepsilon)\phi(n)$, and the probability $P(A_n^{\pm})$ can be estimated using the exponential expression from the theorem[2]. If n_k increases exponentially, i.e., $n_k \sim \theta^k$ for some $\theta > 1$, then the probability $P(A_{n_k}^{\pm})$ behaves according to the order of $k^{-(1\pm\varepsilon)^2}$. This implies that for large k, n_k is asymptotically equivalent to θ^k, where θ is a constant greater than 1. Given this exponential growth rate, n_k increases rapidly with k, as each term in the sequence is effectively multiplied by

[2] As n grows, $\log n$ increases, and hence $(\log n)^{-(1\pm\varepsilon)^2}$ decreases. The idea is that the probability $P(A_n^{\pm})$ is "near" this value in the sense that it follows the same trend; it becomes smaller as n increases, and does so in a comparable way to the decay of $(\log n)^{-(1\pm\varepsilon)^2}$. The sum S_n being above the thresholds $(1\pm\varepsilon)\phi(n)$ behaves similarly to the decay rate of $(\log n)^{-(1\pm\varepsilon)^2}$, due to the relationship given in the theorem for the probability of S_n being greater than any growing threshold $a_n\sqrt{n}$.

the factor θ, which is greater than 1, thereby making each successive term θ times larger than the previous one. [3]. Given the series $\sum k^{-(1\pm\varepsilon)^2}$, its convergence or divergence depends on the sign associated with ε. Specifically, the series converges when the sign is positive and diverges when the sign is negative. This behavior is crucial for applying the Borel–Cantelli lemmas. According to the first Borel–Cantelli lemma, if a series of probabilities converges, then the probability that the corresponding events occur infinitely often is 0. Consequently, there is a zero probability that the event $A^+_{n_k}$, defined as $S_n \geq (1+\varepsilon)\phi(n_k)$, happens for infinitely many values of k.

To justify (15), it's necessary to address that events $A^+_{n_k}$ for $n \neq n_k$ also contribute, suggesting the choice of θ close to 1. Conversely, if events $A^-_{n_k}$ were independent, the second Borel–Cantelli lemma would ensure that $A^-_{n_k}$ occurs infinitely often with probability 1, leading directly to (16). However, due to the dependency among $A^-_{n_k}$ events, a more elaborate argument is needed, involving a larger θ.

To prove (15) we need a results where $M_k = \max\{S_0, S_1, \ldots, S_k\}$ □

Theorem 3.3 (Intermediate result to Prove (15)) — If the X_k are independent simple random variables with mean 0 and variance 1, then for $\alpha \geq \sqrt{2}$.

$$P\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] \leq 2P\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right]. \tag{18}$$

Proof. Let $A_j = \{M_{j-1} < \alpha\sqrt{n} \leq M_j\}$, then

$$P\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] \leq P\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \sum_{j=1}^{n-1} P\left(A_j \cap \left[\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right]\right). \tag{19}$$

Given that the difference $S_n - S_j$ has variance $n - j$, and by leveraging the independence of the X_k's and Chebyshev's inequality, we deduce that the probability term within the summation cannot exceed a certain bound:

$$P\left(A_j \cap \left[\frac{|S_n - S_j|}{\sqrt{n}} > \sqrt{2}\right]\right) = P(A_j)P\left(\frac{|S_n - S_j|}{\sqrt{n}} > \sqrt{2}\right) \leq \frac{1}{2}P(A_j). \tag{20}$$

Since $\bigcup_{j=1}^{n} A_j \subseteq \{M_n \geq \alpha\sqrt{n}\}$,

$$P\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] \leq P\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \frac{1}{2}P\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right]. \tag{21}$$

□

Now we return to the proof of Law of iterated logarithm:

Proof. **Proof of (15).** Given ε, choose θ so that $\theta > 1$ but $\theta^2 < 1 + \varepsilon$. Let $n_k = [\theta^k]$ and $x_k = \Theta(2\log\log n_k)^{1/2}$. By theorems (3.1) and (3.3)

$$P\left[\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right] \leq 2\exp\left[-\frac{1}{2}(x_k - \sqrt{2})^2(1+\xi_k)\right].$$

where $\xi_k \to 0$. In the probability expression featuring the exponential term, the negative part of the exponent grows in proportion to $\theta^2 \log k$ with increasing k. Given that $\theta > 1$, this means the negative component of the exponent eventually becomes larger than $\theta \log k$, indicating that the probability decays at an accelerated rate. Thus:

$$P\left[\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right] \leq \frac{2}{k^\theta}.$$

[3] $n_k \sim \theta^k$ means that the ratio $\frac{n_k}{\theta^k}$ approaches 1 as k goes to infinity. So for large values of k, n_k can be well-approximated by θ^k.

Given $\theta > 1$, the first Borel–Cantelli lemma implies that the event occurs with probability zero

$$M_{n_k} \geq \theta\phi(n_k) \tag{22}$$

for infinitely many k. Suppose that $n_{k-1} < n \leq n_k$ and that

$$S_n > (1+\theta)\phi(n). \tag{23}$$

Now $\phi(n) \geq \phi(n_{k-1}) \sim \theta^{-1/2}\phi(n_k)$; hence, by the choice of θ, $(1+\varepsilon)\phi(n) > \Theta\phi(n_k)$ if k is large enough. Thus for sufficiently large k, (23) implies (22) (if $n_{k-1} < n \leq n_k$), and there is therefore probability 0 that (23) holds for infinitely many n. □

Proof. **Proof of (16)**. Given ε, we choose an integer θ so large that $3\theta^{-1/2} < \varepsilon$. Take $n_k = \theta^k$. Now $n_k - n_{k-1} \to \infty$, and we apply (3.1) applies with $n = n_k - n_{k-1}$ and $a_n = x_k/\sqrt{n_k - n_{k-1}}$, where $x_k = (1-\theta^{-1})\phi(n_k)$. It follows that

$$P\left[S_{n_k} - S_{n_{k-1}} \geq x_k\right] = P\left[S_{n_k - n_{k-1}} \geq x_k\right] = \exp\left[-\frac{1}{2}\frac{x_k^2}{n_k - n_{k-1}}(1+\xi_k)\right],$$

where $\xi_k \to 0$.

For sufficiently large k, the negative exponent in the probability expression asymptotically approaches $(1-\theta^{-1})\log k$, which is less than $\log k$. Consequently, this gives a lower bound for the probability $P\left[S_{n_k} - S_{n_{k-1}} \geq x_k\right]$ that decays slower than k^{-1}. With the independence of the events, the second Borel–Cantelli lemma ensures that almost surely, $S_{n_k} - S_{n_{k-1}}$ will exceed x_k infinitely often.

By applying (15) to the negatives of the X_i's, there is almost surely a bound such that $-S_{n_{k-1}} \leq 2\theta^{-1/2}\phi(n_k)$ for all but finitely many k. These observations imply that S_{n_k} surpasses $x_k - 2\theta^{-1/2}\phi(n_k)$, which, due to the choice of θ, is greater than $(1-\varepsilon)\phi(n_k)$, hence the last inequality holds. □