Section 9

Large Deviations and The Law of The Iterated Logarithm

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

How fast $(X_1 + X_2 + \cdots + X_n)/n$ converges to mean

- After the validation of the strong law, the next question is naturally "how fast $(X_1 + X_2 + \cdots + X_n)/n$ converges to marginal mean?"
- More specifically, we concern that how fast

$$\Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right]$$

converges to zero.

- This concern brings up an interesting research subject—*large deviations*, which uses the technique of *moment generating function* and *Markov's inequality*.

How fast $(X_1 + X_2 + \cdots + X_n)/n$ converges to mean

Definition (Moment generating function) The moment generating function of a random variable X is defined as

$$M(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} e^{tx} dF(x),$$

for all t for which this is finite, where $F(\cdot)$ is the cumulative distribution function (cdf) of the random variable X.

- A general representation for integration w.r.t. a (cumulative) distribution function $F(\cdot)$ is $\int_{\mathcal{X}} \cdot dF(x)$, which is named the *Stieltjes integral*.
- Such a representation can be applied for both discrete support and continuous support. We will use this convention to free the burden of differentiating discrete random variables from continuous random variables.

How moment generating function gets its name?

The Taylor expansion of $\exp\{tx\}$ is equal to $\sum_{k=0}^{\infty} \frac{t^k}{k!} x^k$ for all real x.

Hence, $\exp\{tX\} = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k$, which gives that

$$E\left[\exp\{tX\}\right] = E\left[\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k].$$

if all moments of X exist.

Question: What if some moments of X do not exist? See the next example.

Example If the pdf of X is $1/(x+1)^2$ for $x \geq 0$, and zero, otherwise, then $E[X^k] = \infty$ for $k \geq 1$. In such case, $E[e^{tX}] = \begin{cases} 1 + te^{-t} \int_{-t}^{\infty} s^{-1} e^{-s} ds, & \text{if } t < 0; \\ 1, & \text{if } t = 0; \\ \infty, & \text{if } t > 0. \end{cases}$

But $\sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k] = 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \cdot \infty$ is apparently not a well-defined function.

Therefore, we cannot write $E\left[\exp\{tX\}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k]$.

How moment generating function gets its name?

Lemma ([A29]) If $f(x) = \sum_{k=0}^{\infty} a_k x^k$ converges for every $|x| < r$ for some $r > 0$, then $f(\cdot)$ is differentiable for $|x| < r$, and

$$f'(x) = \frac{df(x)}{dx} = \sum_{k=1}^{\infty} k\, a_k\, x^{k-1}.$$

From the lemma, it is clear that $a_1 = f'(0)$.

If $f'(x) = \sum_{k=1}^{\infty} k\, a_k\, x^{k-1}$ also converges for $|x| < r$ for some $r > 0$, then $a_2 = \frac{1}{2!} f''(0)$.

$\cdots \cdots$

We can repeat the process to obtain all the coefficients, if "convergence-for-some-r" keeps valid.

How moment generating function gets its name?

- Thus, if $E[|X^k|] < A^k$ for some A (which, for example, is valid when X is a bounded random variable), then $E[e^{tX}] = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k]$ (and also every order of its derivatives) converges for $|t| < 1/A$.

- Accordingly, for $k \geq 1$,

$$E[X^k] = M^{(k)}(0) = \left. \frac{d^k M(t)}{dt^k} \right|_{t=0}.$$

This concludes that the moment of X can be obtained by successive differentiation, whence $M(t)$ gets its name.

- **For convenience, we assume throughout (these slides) unless stated otherwise that all moments of the concerned random variables exist, and these moments can be obtained by taking successive derivative of the moment generating function.**

Do moments alone determine the distribution?

Not necessarily, even if all moments exist! We need "positive radius of convergence" for the below power series.

Theorem 30.1 Suppose all moments of X exist. Then if the power series $\sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k]$ has a positive radius of convergence (i.e., convergence for $|t| < r$ for some r), the distribution is uniquely determined by the moments.

Application of the *distribution-determined-by-its-moments* property

Example 9.2 If $X_1, \ldots, X_n$ are i.i.d. with $\Pr[X_n = 1] = 1 - \Pr[X_n = 0] = p$. Determine the distribution of $S_n = X_1 + \cdots + X_n$ by means of Theorem 30.1.

Do moments alone determine the distribution?

Solution: The moment generating function of S_n is:

$$
\begin{aligned}
E[e^{tS_n}] &= E[e^{t(X_1+\cdots+X_n)}] = \prod_{j=1}^{n} E[e^{tX_j}] \quad \text{(by independence)} \\
&= (pe^t + (1-p))^n = \sum_{k=0}^{n} \left[\binom{n}{k} p^k (1-p)^k \right] e^{tk} \quad \text{(by binomial expansion).}
\end{aligned}
$$

Since all moments of S_n exist (by $E[|S_n^k|] \le n^k < \infty$), and

$$
\left| \sum_{k=0}^{\infty} \frac{t^k}{k!} E[S_n^k] \right| \le \sum_{k=0}^{\infty} \frac{|t|^k}{k!} E[|S_n|^k] \le \sum_{k=0}^{\infty} \frac{|t|^k}{k!} n^k = \sum_{k=0}^{\infty} \frac{|tn|^k}{k!} = e^{|tn|} < \infty
$$

for every $t \in \Re$, the distribution of S_n shall equal $\Pr[S_n = k] = \binom{n}{k} p^k (1-p)^k$.□

Definition (Cumulant generating function) The *cumulant generating function* is defined as:

$$C(t) = \log M(t),$$

where $M(t)$ is the moment generating function.

Polynomial approximate to cumulant generating function

- The Taylor expansion of $\log(x+1)$ is $\displaystyle\sum_{v=1}^{\infty} \frac{(-1)^{v+1}}{v} x^v$.

  Hence, if $M(t) = \displaystyle\sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k]$,

$$\begin{aligned} C(t) &= \log M(t) \\ &= \log[(M(t)-1)+1] \\ &= \sum_{v=1}^{\infty} \frac{(-1)^{v+1}}{v} (M(t)-1)^v \\ &= \sum_{v=1}^{\infty} \frac{(-1)^{v+1}}{v} \left(\sum_{k=1}^{\infty} \frac{t^k}{k!} E[X^k] \right)^v \equiv \sum_{i=1}^{\infty} \frac{c_i}{i!} t^i, \end{aligned}$$

  where $\{c_i\}_{i=1}^{\infty}$ are *cumulants* of X.

Cumulant generating function

Equating coefficients leads to

$$\begin{aligned}
C^{(1)}(0) = c_1 &= E[X] \\
C^{(2)}(0) = c_2 &= E[X^2] - E^2[X] = \mathrm{Var}[X] \\
C^{(3)}(0) = c_3 &= E[X^3] - 3E[X][X^2] + 2E^3[X] \\
C^{(4)}(0) = c_4 &= E[X^4] - 4E[X]E[X^3] - 3E^2[X^2] + 12E^2[X]E[X^2] - 6E^4[X] \\
&\vdots
\end{aligned}$$

In case $E[X] = 0$,

$$\begin{aligned}
C^{(1)}(0) = c_1 &= 0 \\
C^{(2)}(0) = c_2 &= E[X^2] = \mathrm{Var}[X] \\
C^{(3)}(0) = c_3 &= E[X^3] \\
C^{(4)}(0) = c_4 &= E[X^4] - 3E^2[X^2] \\
&\vdots
\end{aligned}$$

Properties of cumulant generating function

Property 1 $C_{S_n}(t) = C_{X_1}(t) + C_{X_2}(t) + \cdots + C_{X_n}(t)$ for $S_n = \sum_{j=1}^n X_j$ with independent $\{X_j\}_{j=1}^n$.

Proof: Taking logarithms converts the product relation into additive relation. So, for independent random variables $X_1, X_2, \ldots, X_n$, the cumulant generating function of $S_n = X_1 + X_2 + \cdots + X_n$ equal:

$$C_{S_n}(t) = C_{X_1}(t) + C_{X_2}(t) + \cdots + C_{X_n}(t).$$

□

By this, we can easily confirm that the variance (c_2) of sum of independent samples equal the sum of individual variances (c_2) by taking the 2nd derivatives. This can be extended to any order of cumulants.

Property 2 $C_{X-x}(t) = C_X(t) - t\,x$.

Proof:

$$C_{X-x}(t) = \log E\left[e^{t(X-x)}\right] = \log E\left[e^{tX}\right] - tx = C_X(t) - tx.$$

□

Large deviations

- Observe that

$$
\begin{aligned}
&\Pr\left[\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\right|\geq\varepsilon\right]\\
=\ &\Pr\left[\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\geq\varepsilon\right]+\Pr\left[\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\leq-\varepsilon\right].
\end{aligned}
$$

- By letting $Y=(X_1+X_2+\cdots+X_n)/n$ and $y=\varepsilon+m$ for the first term, and letting $Y=-(X_1+X_2+\cdots+X_n)/n$ and $y=\varepsilon-m$ for the second term, the concerned "probability" becomes the estimate (or find a bound for) $\Pr[Y\geq y]$.

- Can we provide a nice bound to $\Pr[Y\geq y]$ by moments of Y, as moments are "easier" to obtain in practice? For example, for noise in communications, only the first two moments are assumed known usually!

- (One of the) *Answer*(s): Markov's inequality if Y is non-negative random variable.

- What if Y may have negative values.

Large deviations

- *Solution*: Estimate $\Pr[Y \geq y]$ in terms of $\Pr[e^{tY} \geq e^{ty}]$ for any positive t. In other words, since e^{tY} is non-negative, Markov's inequality gives that

$$\Pr[Y \geq y] = \Pr\left[e^{t'Y} \geq e^{t'y}\right] \leq \frac{E\left[(e^{t'Y})^k\right]}{(e^{t'y})^k} = \frac{E\left[e^{tY}\right]}{e^{ty}} = \frac{M_Y(t)}{e^{ty}} = e^{-[ty - C_Y(t)]}.$$

- Therefore, we can provide an upper bound for $\Pr[Y \geq y]$ in terms of the cumulant generating function as:

$$\Pr[Y \geq y] \leq \exp\{-[ty - C_Y(t)]\} = \exp\{C_{Y-y}(t)\} \text{ for } t \geq 0.$$

  - Hence the best exponent is $\inf_{t\geq 0} C_{Y-y}(t)$. For $E[Y] \leq 0$, we will see later that $\inf_{t\geq 0} C_{Y-y}(t) = \inf_{t\in\Re} C_{Y-y}(t)$.
  - Function $I_Y(y) = \sup_{t\in\Re}[ty - C_Y(t)] = -\inf_{t\in\Re} C_{Y-y}(t)$ is called the *large deviation rate function*. We will learn that the large deviation rate function provides the *rate of convergence* very shortly.

Large deviations

It suffices to derive theorems on $\Pr[\bar{Y} \geq 0]$, since we can treat $[Y \geq y]$ as $[\bar{Y} = Y - y \geq 0]$, and still yield the same bound. I.e.,

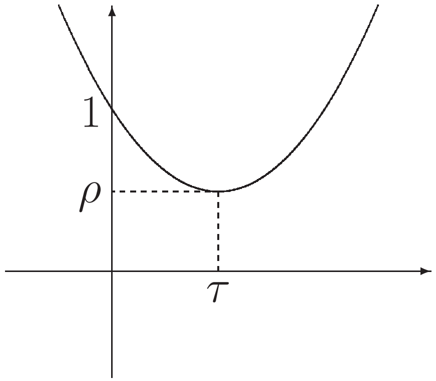
$$\begin{aligned}\Pr[Y \geq y] &= \Pr[(Y-y) \geq 0] \\ &= \Pr[\bar{Y} \geq 0] \\ &\leq \exp\{-[t \cdot 0 - C_{\bar{Y}}(t)]\} \qquad (= M_{\bar{Y}}(t)) \\ &= \exp\{-[t \cdot 0 - C_{(Y-y)}(t)]\} \\ &= \exp\{-[ty - C_Y(t)]\} \text{ for } t \geq 0.\end{aligned}$$

- Although the above formula is always valid, it becomes trivial if $[ty - C_Y(t)] \leq 0$.
- Question is under what condition the bound becomes non-trivial.
  Answer: If $E[\bar{Y}] < 0$, then $\Pr[\bar{Y} \geq 0] \leq \exp\{C_{\bar{Y}}(t)]\} = M_{\bar{Y}}(t) < 1$.

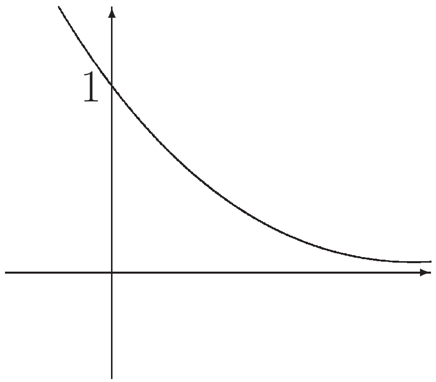
Claim: $E[Y^2] > 0 \Rightarrow (\forall t \in \Re) M''_Y(t) > 0$. *Proof:* Suppose $M''_Y(t) = E[Y^2 e^{tY}] = 0$ for some $t = s$. Then $\Pr[Y^2 e^{sY} \geq 0] = 1$ and $E[Y^2 e^{sY}] = 0$ jointly imply that $\Pr[Y^2 e^{sY} = 0] = 1$. Therefore, $\Pr[Y = 0] = 1$, which contradicts to $E[Y^2] > 0$.

$\Pr[Y \geq 0] \leq M_Y(t)$ under $-\infty < E[Y] < 0$

Observation 1 $M'_Y(0) = E[Y] < 0$ and $M''_Y(0) = E[Y^2] \geq E^2[Y] > 0$ imply that $M_Y(t)$ is **strictly** convex and has minimum at some $t = \tau > 0$, where $\rho = M_Y(\tau) < 1$.



Observation 2 The minimum τ can possibly be infinity, if $\Pr[Y > 0] = 0$.



Example for Observation 2 Suppose $\Pr[Y = -1] = \Pr[Y = 0] = 1/2$. Then $\Pr[Y > 0] = 0$, $M_Y(t) = (e^{-t} + 1)/2$, $\tau = \infty$, and $\rho = \Pr[Y = 0] = 1/2$.

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

- Billingsley's book assumes that $\Pr[Y > 0] > 0$. Indeed, $\Pr[Y > 0] > 0$ implies $M_Y(t) \to \infty$ as $t \to \infty$, which in turns implies $\tau < \infty$, which is actually what desires by the subsequent derivations.
- Notably, $\tau = \infty$ implies $\Pr[Y > 0] = 0$. So $\Pr[Y \geq 0] = \Pr[Y = 0]$, a trivial case of little interest!
- In addition, under $\tau < \infty$ and $-\infty < E[Y] < 0$ (and implicitly, $\tau > 0$), $\rho = M_Y(\tau) > 0$ because if $M_Y(\tau) = E[e^{\tau Y}] = 0$, then $\Pr[e^{\tau Y} = 0] = 1$, which implies $\Pr[Y = -\infty] = 1$, a contradiction to $E[Y] > -\infty$.
- Billingsley's book did not assume that $E[Y] > -\infty$. But this case is actually implicitly included (by assuming *simple* random variable) in the proof.

Example $\Pr[Y = -n] = 3/(\pi^2 n^2)$ for $n \geq 1$, and $\Pr[Y = 0] = \Pr[Y = 1] = 1/4$.

Then $E[Y] = -\infty$ and $M_Y(t) = \dfrac{3}{\pi^2}\displaystyle\sum_{k=1}^{\infty}\frac{e^{-kt}}{k^2} + \frac{1+e^t}{4} = \begin{cases} \infty, & \text{for } t < 0 \\ 1, & \text{for } t = 0 \\ M_Y(t), & \text{for } t > 0 \end{cases}$

Notably, in this example, the moments cannot be determined by taking the derivatives of $M_Y(t)$.

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

- So we have an upper bound $\inf_{t \in \Re} M_Y(t) = M_Y(\tau)$ on $\Pr[Y \geq 0]$ under the condition that all moments exist and the moments can be obtained by taking the successive derivative of the moment generating function.
- Is the upper bound tight? How to obtain a lower bound to $\Pr[Y \geq 0]$?

  Answer: The bound is tight *exponentially*. This can be validated by providing a lower bound to $\Pr[Y \geq 0]$ in terms of the *twisting* distribution technique.

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

Definition (Twisted distribution) The *twisted distribution* of $P_Y(y)$ with twisted parameter t is defined as:

$$dP_{Y^{(t)}}(y) = \frac{e^{ty}dP_Y(y)}{\displaystyle\int_{\mathcal{Y}} e^{ty'}dP_Y(y')} = \frac{e^{ty}dP_Y(y)}{M_Y(t)}.$$

- For discrete random variable, the above equation can be rewritten as:

$$\Pr[Y^{(t)} = y] = \frac{e^{ty}\Pr[Y = y]}{\displaystyle\sum_{y' \in \mathcal{Y}} e^{ty'}\Pr[Y = y']} = \frac{e^{ty}\Pr[Y = y]}{M_Y(t)}.$$

Why twisting a distribution?

Answer: To facilitate a probability estimation.

- In principle,

$$\Pr[Y \geq 0] = \int_0^\infty dP_Y(y) = \int_0^\infty M_Y(t)e^{-ty}dP_{Y^{(t)}}(y) = M_Y(t)\int_0^\infty e^{-ty}dP_{Y^{(t)}}(y).$$

- Suppose $\Pr[Y \geq 0]$ is very, very small (so it is uneasy to make an accurate estimate of it). Then choose t such that $M_Y(t) \ll 1$ immediately gives that

$$\left(\int_0^\infty e^{-ty}dP_{Y^{(t)}}(y)\right) \gg \Pr[Y \geq 0].$$

- Conceptually, measuring a big quantity (e.g., 0.32113 or $\left(\int_0^\infty e^{-ty}dP_{Y^{(t)}}(y)\right)$) is easier than measuring an infinitely small quantity (e.g., $0.0\ldots032113$ or $\Pr[Y \geq 0]$).

- Such a "reversible probability amplifying technique" by moment generating function helps us to make a better estimate on the perhaps very small $\Pr[Y \geq 0]$. (Note: The technique has to be "reversible"; otherwise, we still know little about $\Pr[Y \geq 0]$.)

- *Question*: How to enlarge $\left(\int_0^\infty e^{-ty}dP_{Y^{(t)}}(y)\right)$ to the extreme? Take $t = \tau$ that gives the smallest $M_Y(t)$.

Properties of twisting distribution

- By convention, denote by $Y^{(t)}$ the random variable having $P_{Y^{(t)}}(\cdot)$ as its distribution.

- The moment generating function of $Y^{(t)}$ is:

$$M_{Y^{(t)}}(s) = E\left[e^{sY^{(t)}}\right] = \int_{\mathcal{Y}} e^{sy} dP_{Y^{(t)}}(y) = \int_{\mathcal{Y}} e^{sy} \frac{e^{ty}}{M_Y(t)} dP_Y(y) = \frac{M_Y(s+t)}{M_Y(t)}.$$

- The moments of $Y^{(t)}$ become:

$$E\left[\left(Y^{(t)}\right)^k\right] = \left.\frac{d^k M_{Y^{(t)}}(s)}{ds^k}\right|_{s=0} = \left.\frac{M_Y^{(k)}(s+t)}{M_Y(t)}\right|_{s=0} = \frac{M_Y^{(k)}(t)}{M_Y(t)}.$$

So $E\left[Y^{(\tau)}\right] = \dfrac{M'_Y(\tau)}{M_Y(\tau)} = 0$. In other words, τ—the largest amplifier—has twisted the *mean* to the concerned *margin*, i.e., zero. This enlarges the "desire-to-estimate probability" to the extreme.

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

Now we have an upper bound:

$$\begin{aligned}
\Pr[Y \geq 0] &= M_Y(\tau) \int_0^\infty e^{-\tau y} dP_{Y^{(\tau)}}(y) \\
&\leq M_Y(\tau) \int_0^\infty dP_{Y^{(\tau)}}(y) \quad (\text{since } \tau > 0\ ) \\
&= M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] \\
&\leq M_Y(\tau).
\end{aligned}$$

Is the upper bound close to the true probability? In other words, is the upper bound an overestimate of the true probability?

A straightforward approach to substantiate the tightness of an upper bound is to find a lower bound that is close to the upper bound.

The twisting distribution is especially useful in finding a lower bound.

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

Let $dP_W(y) = \dfrac{dP_{Y^{(\tau)}}(y)}{\Pr\left[Y^{(\tau)} \geq 0\right]}$ be a distribution with support $\{y \in \Re : y \geq 0\}$, and denote its associated random variable by W.

From Slide 9-20, $\Pr[Y^{(\tau)} \geq 0] \geq \Pr[Y \geq 0]/M_Y(\tau)$. As $0 < M_Y(\tau) < 1$, we must have $\Pr[Y^{(\tau)} \geq 0] > 0$ because if it were not true, $\Pr[Y \geq 0] = 0$ leads to $\tau = \infty$.

In fact, $\Pr[Y \geq 0] = 0$ already gives us what we demand to derive, i.e., the probability of $\Pr[Y \geq 0]$! So the subsequent derivation is unnecessary for such a trivial case.

$$\begin{aligned}
\Pr[Y \geq 0] &= M_Y(\tau) \int_0^\infty e^{-\tau y} dP_{Y^{(\tau)}}(y) \\
&= M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] \int_0^\infty e^{-\tau y} \frac{dP_{Y^{(\tau)}}(y)}{\Pr\left[Y^{(\tau)} \geq 0\right]} \\
&= M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] \int_0^\infty e^{-\tau w} dP_W(w) \\
&= M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] E[e^{-\tau W}].
\end{aligned}$$

(Jensen's Inequality) For a convex function $\varphi(\cdot)$, $\varphi(E[X]) \leq E[\varphi(X)]$.

By Jensen's inequality with $\varphi(x) = e^{-\tau x}$,

$$e^{-\tau E[W]} = \varphi(E[W]) \leq E[\varphi(W)] = E\left[e^{-\tau W}\right].$$

Bound under $-\infty < E[Y] < 0$ and $\tau < \infty$

This gives that:

$$\Pr[Y \geq 0] \;\geq\; M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] e^{-\tau E[W]}.$$

Observe that

$$\begin{aligned}
E[W] &= \int_0^{\infty} w\, dP_W(w) \\
&= \int_0^{\infty} y \frac{dP_{Y^{(\tau)}}(y)}{\Pr\left[Y^{(\tau)} \geq 0\right]} \\
&\leq \frac{1}{\Pr\left[Y^{(\tau)} \geq 0\right]} \int_{-\infty}^{\infty} |y|\, dP_{Y^{(\tau)}}(y) \\
&= \frac{1}{\Pr\left[Y^{(\tau)} \geq 0\right]} E\left[\left|Y^{(\tau)}\right|\right] \\
&\leq \frac{1}{\Pr\left[Y^{(\tau)} \geq 0\right]} E^{1/2}\left[\left(Y^{(\tau)}\right)^2\right] \quad \text{(by Lyapounov's ineq.)}
\end{aligned}$$

We can then conclude that if $-\infty < E[Y] < 0$ and $\tau < \infty$, then

$$\underline{M_Y(\tau)} \geq \Pr[Y \geq 0] \geq \underline{M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] \exp\left\{-\frac{\tau}{\Pr\left[Y^{(\tau)} \geq 0\right]} E^{1/2}\left[\left(Y^{(\tau)}\right)^2\right]\right\}}.$$

A lower bound to $\Pr[U \geq 0]$ for zero mean U

To replace $\Pr[Y^{(\tau)} \geq 0]$ in the previous inequality by moments of $Y^{(\tau)}$, we need a lower bound to it.

Theorem 9.2 If $E[U] = 0$ and $E[U^2] > 0$, then

$$\Pr[U \geq 0] \geq \frac{E^2[U^2]}{4E[U^4]}.$$

(Schwarz's inequality) $E[|XY|] \leq E^{1/2}[X^2]E^{1/2}[Y^2]$

(Hölder's inequality) $E[|XY|] \leq E^{1/p}[|X|^p]E^{1/q}[|Y|^q]$ for $p > 1$, $q > 1$ and $1/p + 1/q = 1$

Proof:

- Let $U^+ = U \cdot I_{[U \geq 0]}$ and $U^- = -U \cdot I_{[U<0]}$, where $I_{[\cdot]}$ is an indicator random variable that equals 1 when the event is true, and 0 when the event is false.

  By their definitions, U^+ and U^- are both non-negative.

- Then by Schwarz's inequality,

$$E\left[(U^+)^2\right] = E\left[U^2 \cdot I_{[U\geq 0]}\right] \leq E^{1/2}\left[U^4\right] E^{1/2}\left[I^2_{[U\geq 0]}\right] = E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right],$$

and by Hölder's inequality,

$$\begin{aligned} E\left[(U^-)^2\right] &= E\left[\left|U^-\right|^{2/3}\left|U^-\right|^{4/3}\right] \\ &\leq E^{1/(3/2)}\left[\left(\left|U^-\right|^{2/3}\right)^{3/2}\right] E^{1/3}\left[\left(\left|U^-\right|^{4/3}\right)^{3}\right] \\ &= E^{2/3}\left[\left|U^-\right|\right] E^{1/3}\left[\left|U^-\right|^4\right] \\ &\leq E^{2/3}\left[U^-\right] E^{1/3}\left[U^4\right]. \end{aligned}$$

Also by Hölder's inequality, together with $E[U] = 0$ and $U = U^+ - U^-$,

$$\begin{aligned} E[U^-] &= E[U^+] \\ &= E[U \cdot I_{[U \geq 0]}] \\ &= E[|U \cdot I_{[U \geq 0]}|] \\ &\leq E^{1/4}[U^4]\, E^{3/4}[I_{[U \geq 0]}^{4/3}] \\ &= E^{1/4}[U^4]\, E^{3/4}[I_{[U \geq 0]}]. \end{aligned}$$

A lower bound to $\Pr[U \geq 0]$ for zero mean U

Hence,

$$\begin{aligned} E[U^2] &= E[(U^+)^2] + E[(U^-)^2] \\ &\leq E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right] + E^{2/3}\left[U^-\right]\ E^{1/3}\left[U^4\right] \\ &\leq E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right] + \left(E^{1/4}[U^4]\, E^{3/4}[I_{[U\geq 0]}]\right)^{2/3}\ E^{1/3}\left[U^4\right] \\ &= E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right] + E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right] \\ &= 2E^{1/2}\left[U^4\right] E^{1/2}\left[I_{[U\geq 0]}\right] \\ &= 2E^{1/2}\left[U^4\right] (\Pr[U \geq 0])^{1/2}, \end{aligned}$$

which immediately validates the theorem. □

A lower bound to $\Pr[U \geq 0]$ for zero mean U

$$E[(Y^{(\tau)})^2] > 0 \Rightarrow \Pr[Y^{(\tau)} \geq 0] \geq \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]}$$

$E[(Y^{(\tau)})^2] = 0 \Rightarrow \Pr[Y^{(\tau)} = 0] = 1 \Rightarrow \Pr[Y = 0] = 1$ because $Y^{(\tau)}$ has the same support as $Y \Rightarrow E[Y] = 0$, a contradiction to the assumption that $-\infty < E[Y] < 0$.

$$\begin{aligned}
\Rightarrow M_Y(\tau) \geq \Pr[Y \geq 0] &\geq M_Y(\tau) \cdot \Pr\left[Y^{(\tau)} \geq 0\right] \exp\left\{-\frac{\tau}{\Pr\left[Y^{(\tau)} \geq 0\right]} E^2[(Y^{(\tau)})^2]\right\} \\
&\geq M_Y(\tau) \cdot \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} \exp\left\{-\frac{\tau}{\frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]}} E^{1/2}[(Y^{(\tau)})^2]\right\} \\
&= M_Y(\tau) \cdot \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} \exp\left\{-\frac{4\tau E[(Y^{(\tau)})^4]}{E^{3/2}[(Y^{(\tau)})^2]}\right\}
\end{aligned}$$

where

$$\begin{cases} E[(Y^{(\tau)})^2] = C''_{Y^{(\tau)}}(0) = C''_Y(\tau) \\ E[(Y^{(\tau)})^4] = C^{(4)}_{Y^{(\tau)}}(0) + 3E^2[(Y^{(\tau)})^2] = C^{(4)}_Y(\tau) + 3\left(C''_Y(\tau)\right)^2 \end{cases}$$

Chernoff's Theorem

- Recall that what we concern is:

$$\begin{aligned}
&\Pr\left[\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\right|\geq\varepsilon\right]\\
&= \Pr\left[\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\geq\varepsilon\right]\\
&\quad+\Pr\left[\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\leq-\varepsilon\right]\\
&= \Pr\left[(X_1-m-\varepsilon)+(X_2-m-\varepsilon)+\cdots+(X_n-m-\varepsilon)\geq 0\right]\\
&\quad+\Pr\left[(m-X_1-\varepsilon)+(m-X_2-\varepsilon)+\cdots+(m-X_n-\varepsilon)\geq 0\right]\\
&= \Pr[Y\geq 0]+\Pr[\hat{Y}\geq 0],
\end{aligned}$$

where

$$Y=(X_1-m-\varepsilon)+(X_2-m-\varepsilon)+\cdots+(X_n-m-\varepsilon)$$

and

$$\hat{Y}=(m-X_1-\varepsilon)+(m-X_2-\varepsilon)+\cdots+(m-X_n-\varepsilon).$$

- Assume $\{X_i\}_{i=1}^n$ are i.i.d.

Chernoff's Theorem: Upper bound

$$M_Y(\tau) \geq \Pr[Y \geq 0] \geq M_Y(\tau) \cdot \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} \exp\left\{ -\frac{4\tau E[(Y^{(\tau)})^4]}{E^{3/2}[(Y^{(\tau)})^2]} \right\}.$$

- $-\infty < E[Y] = n(E[X] - m - \varepsilon) = -n\varepsilon < 0.$

- $M_Y(t) = \prod_{i=1}^{n} M_{(X_i - m - \varepsilon)}(t) = M^n_{(X-m-\varepsilon)}(t)$

  $\Rightarrow M'_Y(t) = nM'_{(X-m-\varepsilon)}(t) M^{n-1}_{(X-m-\varepsilon)}(t)$

  $\Rightarrow \tau$ satisfies $M'_{(X-m-\varepsilon)}(\tau) = 0$, which is **assumed finite**.

  Or equivalently, we can say τ satisfies $C'_{(X-m-\varepsilon)}(\tau) = 0$.

  Notably, τ is independent of n.

- Let $\rho = M_{(X-m-\varepsilon)}(\tau)$. Since ρ is the minimizer for $M_{(X-m-\varepsilon)}(t)$ that has negative mean $-\varepsilon$ and $\tau < \infty$, we have $0 < \rho < 1$.

- We then obtain:

$$\rho^n \geq \Pr[Y \geq 0]$$

Chernoff's Theorem: Lower bound

$$M_Y(\tau) \geq \Pr[Y \geq 0] \geq M_Y(\tau) \cdot \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} \exp\left\{-\frac{4\tau E[(Y^{(\tau)})^4]}{E^{3/2}[(Y^{(\tau)})^2]}\right\}.$$

- $M_Y(t) = M^n_{(X-m-\varepsilon)}(t)$
  $\Rightarrow \underline{M_Y(\tau)} = M^n_{(X-m-\varepsilon)}(\tau) = \underline{\rho^n}$
- $C_Y(t) = nC_{(X-m-\varepsilon)}(t) = n[C_X(t) - (m+\varepsilon)t]$
- $E[(Y^{(\tau)})^2] = C''_{Y^{(\tau)}}(0) = C''_Y(\tau) = nC''_X(\tau) = n\lambda_2,$
  where $\lambda_2 = C''_X(\tau)$.
- $E[(Y^{(\tau)})^4] = C^{(4)}_{Y^{(\tau)}}(0) + 3E^2[(Y^{(\tau)})^2] = C^{(4)}_Y(\tau) + 3n^2\lambda_2^2$
  $= nC^{(4)}_X(\tau) + 3n^2\lambda_2^2 = n\lambda_4 + 3n^2\lambda_2^2,$
  where $\lambda_4 = C^{(4)}_X(\tau)$.
- $\Rightarrow \dfrac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} = \dfrac{E^2[(Y^{(\tau)})^2]/n^2}{4E[(Y^{(\tau)})^4]/n^2} = \dfrac{\lambda_2^2}{4(\lambda_4/n + 3\lambda_2^2)} = \dfrac{1}{12 + 4\lambda_4/(\lambda_2^2 n)}$

Chernoff's Theorem: Lower bound

- $\dfrac{4\tau E[(Y^{(\tau)})^4]}{E^{3/2}[(Y^{(\tau)})^2]} = \dfrac{4\tau[n\lambda_4 + 3n^2\lambda_2^2]}{n^{3/2}\lambda_2^{3/2}} = 4\tau\lambda_2^{1/2}\left[3 + \lambda_4/(\lambda_2^2 n)\right]\sqrt{n}$

$$\begin{aligned}\Pr[Y \ge 0] &\ge M_Y(\tau) \cdot \frac{E^2[(Y^{(\tau)})^2]}{4E[(Y^{(\tau)})^4]} \exp\left\{-\frac{4\tau E[(Y^{(\tau)})^4]}{E^{3/2}[(Y^{(\tau)})^2]}\right\} \\ &= \rho^n \cdot \frac{1}{12 + 4\lambda_4/(\lambda_2^2 n)} \exp\left\{-4\tau\lambda_2^{1/2}\left[3 + \lambda_4/(\lambda_2^2 n)\right]\sqrt{n}\right\}\end{aligned}$$

- As proved in Slide 9-26, $\lambda_2 = E[(Y^{(\tau)})^2]/n > 0$.
- $n\lambda_4 + 3n^2\lambda_2^2 = E[(Y^{(\tau)})^4] \ge E^2[(Y^{(\tau)})^2] = n^2\lambda_2^2$ implies $\lambda_4 \ge -2n\lambda_2^2$. However, λ_4 is not necessarily positive.

Chernoff's Theorem

This concludes:

$$\rho^n \geq \Pr[Y \geq 0] \geq \rho^n \cdot \frac{1}{12 + 4\lambda_4/(\lambda_2^2 n)} \exp\left\{-4\tau\lambda_2^{1/2}\left[3 + \lambda_4/(\lambda_2^2 n)\right]\sqrt{n}\right\},$$

where

$$\begin{cases} \tau \text{ is the unique, positive, finite solution to } C'_X(\tau) = (m+\varepsilon), \text{ independent of } n \\ \log\rho = C_X(\tau) - (m+\varepsilon)\tau \\ \lambda_2 = C''_X(\tau) \\ \lambda_4 = C_X^{(4)}(\tau) \end{cases}$$

Both the upper and lower bounds can be improved up to $\frac{C}{\sqrt{n}}\rho^n$ (See J. A. Fill and M. J. Wichura, "The convergence rate for the strong law of large numbers: General lattice distributions," Probab. Th. Rel. Fields, 81:189-212, 1989). But the current bounds are sufficient for our present goal.

How about $\Pr[\hat{Y} \geq 0]$? The same technique can be applied.

Chernoff's Theorem: Upper bound

$$M_{\hat{Y}}(\hat{\tau}) \ge \Pr[\hat{Y} \ge 0] \ge M_{\hat{Y}}(\hat{\tau}) \cdot \frac{E^2[(\hat{Y}^{(\hat{\tau})})^2]}{4E[(\hat{Y}^{(\hat{\tau})})^4]} \exp\left\{ -\frac{4\hat{\tau}E[(\hat{Y}^{(\hat{\tau})})^4]}{E^{3/2}[(\hat{Y}^{(\hat{\tau})})^2]} \right\}.$$

- $-\infty < E[\hat{Y}] = n(m - E[X] - \varepsilon) = -n\varepsilon < 0.$
- $M_{\hat{Y}}(t) = \prod_{i=1}^{n} M_{(m-X_i-\varepsilon)}(t) = M^n_{(m-X-\varepsilon)}(t)$

  $\Rightarrow M'_{\hat{Y}}(t) = nM'_{(m-X-\varepsilon)}(t)M^{n-1}_{(m-X-\varepsilon)}(t)$

  $\Rightarrow \hat{\tau}$ satisfies $M'_{(m-X-\varepsilon)}(\hat{\tau}) = 0$, which is **assumed finite**.

  Or equivalently, we can say $\hat{\tau}$ satisfies $C'_{(m-X-\varepsilon)}(\hat{\tau}) = 0$.

  Notably, $\hat{\tau}$ is independent of n.
- Let $\hat{\rho} = M_{(m-X-\varepsilon)}(\hat{\tau})$. Since $\hat{\rho}$ is the minimizer for $M_{(m-X-\varepsilon)}(t)$ that has negative mean $-\varepsilon$ and $\hat{\tau} < \infty$, $0 < \hat{\rho} < 1$.
- We then obtain:

$$\hat{\rho}^n \ge \Pr[\hat{Y} \ge 0]$$

Chernoff's Theorem: Lower bound

$$M_{\hat{Y}}(\hat{\tau}) \geq \Pr[\hat{Y} \geq 0] \geq M_{\hat{Y}}(\hat{\tau}) \cdot \frac{E^2[(\hat{Y}^{(\hat{\tau})})^2]}{4E[(\hat{Y}^{(\hat{\tau})})^4]} \exp\left\{-\frac{4\hat{\tau}E[(\hat{Y}^{(\hat{\tau})})^4]}{E^{3/2}[(\hat{Y}^{(\hat{\tau})})^2]}\right\}.$$

- $M_{\hat{Y}}(t) = M^n_{(m-X-\varepsilon)}(t)$

  $\Rightarrow \underline{M_{\hat{Y}}(\hat{\tau})} = M^n_{(m-X-\varepsilon)}(\hat{\tau}) = \underline{\hat{\rho}^n}$
- $C_{\hat{Y}}(t) = nC_{(m-X-\varepsilon)}(t) = n[C_X(-t) + (m-\varepsilon)t]$
- $E[(\hat{Y}^{(\hat{\tau})})^2] = C''_{\hat{Y}^{(\hat{\tau})}}(0) = C''_{\hat{Y}}(\hat{\tau}) = nC''_X(-\hat{\tau}) = n\hat{\lambda}_2,$

  where $\hat{\lambda}_2 = C''_X(-\hat{\tau})$.
- $E[(\hat{Y}^{(\hat{\tau})})^4] = C^{(4)}_{\hat{Y}^{(\hat{\tau})}}(0) + 3E^2[(\hat{Y}^{(\hat{\tau})})^2] = C^{(4)}_{\hat{Y}}(\hat{\tau}) + 3n^2\hat{\lambda}_2^2$

  $$= nC^{(4)}_X(-\hat{\tau}) + 3n^2\hat{\lambda}_2^2 = n\hat{\lambda}_4 + 3n^2\hat{\lambda}_2^2,$$

  where $\hat{\lambda}_4 = C^{(4)}_X(-\hat{\tau})$.
- $\Rightarrow \dfrac{E^2[(\hat{Y}^{(\hat{\tau})})^2]}{4E[(\hat{Y}^{(\hat{\tau})})^4]} = \dfrac{E^2[(\hat{Y}^{(\hat{\tau})})^2]/n^2}{4E[(\hat{Y}^{(\hat{\tau})})^4]/n^2} = \dfrac{1}{12 + 4\hat{\lambda}_4/(\hat{\lambda}_2^2 n)}$

Chernoff's Theorem: Lower bound

- $\dfrac{4\hat{\tau}E[(\hat{Y}^{(\hat{\tau})})^4]}{E^{3/2}[(\hat{Y}^{(\hat{\tau})})^2]} = \dfrac{4\hat{\tau}[n\hat{\lambda}_4 + 3n^2\hat{\lambda}_2^2]}{n^{3/2}\hat{\lambda}_2^{3/2}} = 4\hat{\tau}\hat{\lambda}_2^{1/2}\left[3 + \hat{\lambda}_4/(\hat{\lambda}_2^2 n)\right]\sqrt{n}$

$$\begin{aligned}\Pr[\hat{Y} \ge 0] &\ge M_{\hat{Y}}(\hat{\tau}) \cdot \frac{E^2[(\hat{Y}^{(\hat{\tau})})^2]}{4E[(\hat{Y}^{(\hat{\tau})})^4]} \exp\left\{-\frac{4\hat{\tau}E[(\hat{Y}^{(\hat{\tau})})^4]}{E^{3/2}[(\hat{Y}^{(\hat{\tau})})^2]}\right\} \\ &= \hat{\rho}^n \cdot \frac{1}{12 + 4\hat{\lambda}_4/(\hat{\lambda}_2^2 n)} \exp\left\{-4\hat{\tau}\hat{\lambda}_2^{1/2}\left[3 + \hat{\lambda}_4/(\hat{\lambda}_2^2 n)\right]\sqrt{n}\right\}\end{aligned}$$

- As proved in Slide 9-26, $\hat{\lambda}_2 = E[(\hat{Y}^{(\hat{\tau})})^2]/n > 0$.
- $n\hat{\lambda}_4 + 3n^2\hat{\lambda}_2^2 = E[(\hat{Y}^{(\hat{\tau})})^4] \ge E^2[(\hat{Y}^{(\hat{\tau})})^2] = n^2\hat{\lambda}_2^2$ implies $\hat{\lambda}_4 \ge -2n\hat{\lambda}_2^2$. However, $\hat{\lambda}_4$ is not necessarily positive.

This concludes to:

$$\hat{\rho}^n \geq \Pr[\hat{Y} \geq 0] \geq \hat{\rho}^n \cdot \frac{1}{12 + 4\hat{\lambda}_4/(\hat{\lambda}_2^2 n)} \exp\left\{-4\hat{\tau}\hat{\lambda}_2^{1/2}\left[3 + \hat{\lambda}_4/(\hat{\lambda}_2^2 n)\right]\sqrt{n}\right\},$$

where

$$\begin{cases} \hat{\tau} \text{ is the unique, positive, finite solution to } C_X'(-\hat{\tau}) = (m - \varepsilon), \text{ independent of } n \\ \log \hat{\rho} = C_X(-\hat{\tau}) + (m - \varepsilon)\hat{\tau} \\ \hat{\lambda}_2 = C_X''(-\hat{\tau}) \\ \hat{\lambda}_4 = C_X^{(4)}(-\hat{\tau}) \end{cases}$$

Both the upper and lower bounds can be improved up to $\frac{C}{\sqrt{n}}\rho^n$ (See J. A. Fill and M. J. Wichura, "The convergence rate for the strong law of large numbers: General lattice distributions," Probab. Th. Rel. Fields, 81:189-212, 1989). But the current bounds are sufficient for our present goal.

Chernoff's Theorem

To summarize,

$$\rho^n + \hat{\rho}^n \geq \Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right]$$
$$\geq \rho^n \cdot \frac{1}{12 + 4\lambda_4/(\lambda_2^2 n)} e^{-4\tau\left[3+\lambda_4/(\lambda_2^2 n)\right]\sqrt{\lambda_2 n}} + \hat{\rho}^n \cdot \frac{1}{12 + 4\hat{\lambda}_4/(\hat{\lambda}_2^2 n)} e^{-4\hat{\tau}\left[3+\hat{\lambda}_4/(\hat{\lambda}_2^2 n)\right]\sqrt{\hat{\lambda}_2 n}}$$

where

$$\begin{cases} \rho = \inf_{t\in\Re} M_{(X-m-\varepsilon)}(t) = \inf_{t\in\Re} e^{-t\varepsilon} e^{-tm} M_X(t) \\ \qquad = \exp\left\{-\sup_{t\in\Re}[t(m+\varepsilon) - C_X(t)]\right\} = \exp\{-I_X(m+\varepsilon)\} \\ \hat{\rho} = \inf_{t\in\Re} M_{(m-X-\varepsilon)}(t) = \inf_{t\in\Re} e^{-t\varepsilon} e^{tm} M_X(-t) \quad \text{(Set } s = -t\text{)} \\ \quad = \inf_{s\in\Re} e^{s\varepsilon} e^{-sm} M_X(s) = \exp\left\{-\sup_{s\in\Re}[s(m-\varepsilon) - C_X(s)]\right\} = \exp\{-I_X(m-\varepsilon)\} \end{cases}$$

where $I_X(x) = \sup_{t\in\Re}[tx - C_X(t)]$ is the large deviation rate function.

Chernoff's Theorem

Observation How fast

$$\Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| > \varepsilon\right]$$

converges to zero, if X_1, X_2, ..., X_n are i.i.d.?

Answer Exactly exponentially fast (we have both upper and lower bounds). And the exponential rate can be determined by the *large deviation rate function* of the marginal X.

Example 1

Suppose we wish to estimate the variance σ^2 of a zero-mean Gaussian noise by successive observations N_1, N_2, …, N_n. Then

$$\Pr\left\{\left|\frac{N_1^2+N_2^2+\cdots+N_n^2}{n}-\sigma^2\right|>\varepsilon\right\}\le\rho^n+\hat{\rho}^n,$$

where $\rho=\exp\left\{-I_{N^2}(\sigma^2+\varepsilon)\right\}$ and $\hat{\rho}=\exp\left\{-I_{N^2}(\sigma^2-\varepsilon)\right\}$.

$$\begin{aligned}
M_{N^2}(t) &= E[e^{tN^2}]=\int_{-\infty}^{\infty}e^{tu^2}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-u^2/(2\sigma^2)}du=\begin{cases}\dfrac{1}{\sqrt{1-2\sigma^2 t}}, & \text{if } t<1/(2\sigma^2);\\ \infty, & \text{otherwise.}\end{cases}\\
\Rightarrow C_{N^2}(t) &= \begin{cases}-\dfrac{1}{2}\log(1-2\sigma^2 t), & \text{if } t<1/(2\sigma^2);\\ \infty, & \text{otherwise.}\end{cases}\\
\Rightarrow I_{N^2}(x) &= \sup_{t\in\Re}[tx-C_{N^2}(t)]=\sup_{t<1/(2\sigma^2)}\left[tx+\frac{1}{2}\log(1-2\sigma^2 t)\right]\\
&= \begin{cases}\infty, & \text{if } x\le 0;\\ \left(\dfrac{1}{2\sigma^2}-\dfrac{1}{2x}\right)x+\dfrac{1}{2}\log\left(1-2\sigma^2\left(\dfrac{1}{2\sigma^2}-\dfrac{1}{2x}\right)\right), & \text{if } x>0\end{cases}\\
&= \begin{cases}\infty, & \text{if } x\le 0;\\ \dfrac{1}{2}\left(\dfrac{x}{\sigma^2}-1\right)-\dfrac{1}{2}\log\left(\dfrac{x}{\sigma^2}\right), & \text{if } x>0\end{cases}
\end{aligned}$$

Example 1

Hence, if $\varepsilon = 0.01\sigma^2$, then

$$\begin{cases} -\log(\rho) = I_{N^2}(\sigma^2+\varepsilon) = \frac{1}{2}(0.01) - \frac{1}{2}\log(1+0.01) = 2.48346 \times 10^{-5} \\ -\log(\hat{\rho}) = I_{N^2}(\sigma^2-\varepsilon) = -\frac{1}{2}(0.01) - \frac{1}{2}\log(1-0.01) = 2.51679 \times 10^{-5} \end{cases}$$

So, if we desire that

$$\Pr\left\{\left|\frac{N_1^2+N_2^2+\cdots+N_n^2}{n} - \sigma^2\right| > 0.01\sigma^2\right\} \le \rho^n + \hat{\rho}^n \le 10^{-5} + 10^{-5},$$

it is safer to have the sample number satisfying:

$$n \ge \max\left\{\frac{\log(10^{-5})}{\log(\rho)}, \frac{\log(10^{-5})}{\log(\hat{\rho})}\right\} = \max\{463583, 457444\} = 463583.$$

□

Observations

- When the required degree of accuracy is higher (i.e., ε smaller), more samples are needed!
- The above required sample number has nothing to do with σ^2.

Example 1

$$C_{N^2}(t) = \begin{cases} -\dfrac{1}{2}\log(1-2\sigma^2 t), & \text{if } t < 1/(2\sigma^2); \\ \infty, & \text{otherwise} \end{cases}$$

$$\begin{aligned}
\Rightarrow \tau &= \frac{\varepsilon}{2\sigma^2(\sigma^2+\varepsilon)} \text{ and } \rho = \frac{1}{\sigma}(\sigma^2+\varepsilon)^{1/2} e^{-\varepsilon/(2\sigma^2)} \\
\Rightarrow \lambda_2 &= C''_{N^2}(\tau) = 2(\sigma^2+\varepsilon)^2 \\
\Rightarrow \lambda_4 &= C^{(4)}_{N^2}(\tau) = 48(\sigma^2+\varepsilon)^4 \\
\Rightarrow \hat{\tau} &= \frac{\varepsilon}{2\sigma^2(\sigma^2-\varepsilon)} \text{ and } \hat{\rho} = \frac{1}{\sigma}(\sigma^2-\varepsilon)^{1/2} e^{\varepsilon/(2\sigma^2)} \\
\Rightarrow \hat{\lambda}_2 &= C''_{N^2}(-\hat{\tau}) = 2(\sigma^2-\varepsilon)^2 \\
\Rightarrow \hat{\lambda}_4 &= C^{(4)}_{N^2}(-\hat{\tau}) = 48(\sigma^2-\varepsilon)^4 \\
\Rightarrow \lambda_4/\lambda_2^2 = \hat{\lambda}_4/\hat{\lambda}_2^2 &= 12 \text{ and } \tau\lambda_2^{1/2} = \hat{\tau}\hat{\lambda}_2^{1/2} = \varepsilon/(\sqrt{2}\sigma^2)
\end{aligned}$$

$$\begin{aligned}
\Pr\left\{\left|\frac{N_1^2+N_2^2+\cdots+N_n^2}{n} - \sigma^2\right| > 0.01\sigma^2\right\} &\ge (\rho^n + \hat{\rho}^n)\,\frac{e^{-0.06(1+4/n)\sqrt{2n}}}{12(1+4/n)} \\
&= 6.64514 \times 10^{-27}\,(\rho^n + \hat{\rho}^n) \text{ for } n = 463853 \text{ and } \varepsilon = 0.01\sigma^2.
\end{aligned}$$

Example 2

Suppose we wish to estimate the head-appearing probability of a coin flip by successive trials X_1, X_2, ..., X_n, where $X_i = 1$ represents a head appearance. Then

$$\Pr\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - p\right| > \varepsilon\right\} \le \rho^n + \hat{\rho}^n,$$

where $\rho = \exp\{-I_X(p+\varepsilon)\}$ and $\hat{\rho} = \exp\{-I_X(p-\varepsilon)\}$.

$$\begin{aligned}
M_X(t) &= E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} dP_X(x) = e^t p + 1 - p \\
\Rightarrow C_X(t) &= \log(e^t p + 1 - p) \\
\Rightarrow I_X(x) &= \sup_{t\in\Re}[tx - C_X(t)] = \sup_{t\in\Re}\left[tx - \log(e^t p + 1 - p)\right] \\
&= \begin{cases} x\log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right), & \text{if } 0 < x < 1; \\ -\log(p), & \text{if } x = 1; \\ -\log(1-p), & \text{if } x = 0; \\ \infty, & \text{otherwise} \end{cases}
\end{aligned}$$

Hence, if $p = 0.1$ and $\varepsilon = 0.01p = 10^{-3}$, then

$$\begin{cases} -\log(\rho) = I_X(p+\varepsilon) = 0.101\log\frac{0.101}{0.1} + 0.899\log\frac{0.899}{0.9} = 5.53918 \times 10^{-6} \\ -\log(\hat{\rho}) = I_X(p-\varepsilon) = 0.099\log\frac{0.099}{0.1} + 0.901\log\frac{0.901}{0.9} = 5.5721 \times 10^{-6} \end{cases}$$

Example 2

So, if we desire that

$$\Pr\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n}-p\right|>0.01p\right\}\le\rho^n+\hat{\rho}^n\le 10^{-5}+10^{-5},$$

it is safer to have the sample number satisfying:

$$n\ge\max\left\{\frac{\log(10^{-5})}{\log(\rho)},\frac{\log(10^{-5})}{\log(\hat{\rho})}\right\}=\max\{2.07845\times10^6,2.06617\times10^6\}=2078450.$$

□

Observation

- When $p=0.5$ and $\varepsilon=0.005$, $-\log(\rho)=-\log(\hat{\rho})=5.00008\times10^{-5}$. Hence, the required number of samples becomes 230255. Accordingly, it is easier to "ensure" a fair coin than a biased coin.

- From the two examples, you shall learn that what the law of large numbers concern is really a **large** number.

Example 2

$$C_X(t) = \log(e^t p + 1 - p)$$

$$\begin{aligned}
\Rightarrow \tau &= \log \frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)} \text{ and } \rho = \left(\frac{1-p}{1-p-\varepsilon}\right)^{1-p-\varepsilon} \left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon} \\
\Rightarrow \lambda_2 &= C''_X(\tau) = (p+\varepsilon)(1-p-\varepsilon) \\
\Rightarrow \lambda_4 &= C^{(4)}_X(\tau) = (p+\varepsilon)(1-p-\varepsilon)[1-6(p+\varepsilon)+6(p+\varepsilon)^2] \\
\Rightarrow \hat{\tau} &= \log \frac{p(1-p+\varepsilon)}{(1-p)(p-\varepsilon)} \text{ and } \hat{\rho} = \left(\frac{1-p}{1-p+\varepsilon}\right)^{1-p+\varepsilon} \left(\frac{p}{p-\varepsilon}\right)^{p-\varepsilon} \\
\Rightarrow \hat{\lambda}_2 &= C''_X(-\hat{\tau}) = (p-\varepsilon)(1-p+\varepsilon) \\
\Rightarrow \hat{\lambda}_4 &= C^{(4)}_X(-\hat{\tau}) = (p-\varepsilon)(1-p+\varepsilon)[1-6(p-\varepsilon)+6(p-\varepsilon)^2]
\end{aligned}$$

$$\Pr\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - p\right| > 0.01p\right\} \geq (7.52188 \times 10^{-27})\rho^n + (7.5219 \times 10^{-27})\hat{\rho}^n$$

for $n = 2078450, p = 0.1, \varepsilon = 0.01p, \tau = 0.0110621, \lambda_2 = 0.090799, \lambda_4 = 0.0413322$
$\hat{\tau} = 0.0111608, \hat{\lambda}_2 = 0.089199, \hat{\lambda}_4 = 0.0414602$

Example 2

$$\Pr\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n}-p\right|>0.01p\right\}\geq(8.20011\times 10^{-27})\rho^n+(8.20011\times 10^{-27})\hat{\rho}^n$$

for $n = 230255, p = 0.5, \varepsilon = 0.01p, \tau = 0.0200007, \lambda_2 = 0.249975, \lambda_4 = -0.12495$

$\hat{\tau} = 0.0200007, \hat{\lambda}_2 = 0.249975, \hat{\lambda}_4 = -0.12495$

Strong law of large number revisited

Theorem If $\sum_{n=1}^{\infty} \Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right] < \infty$ for any $\varepsilon > 0$ arbitrarily small, then the strong law holds.

Theorem If $\{X_j\}_{j=1}^{\infty}$ are i.i.d., and all moments of X_i exist and can be determined by taking the derivatives of the moment generating function, then

$$\sum_{n=1}^{\infty} \Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right] \leq \sum_{n=1}^{\infty}(\rho^n + \hat{\rho}^n) < \infty$$

for any $\varepsilon > 0$ arbitrarily small; hence, the strong law holds.

Proof: Choose $0 < \tau < \infty$ such that $M_{(X-m-\varepsilon))}(\tau) < 1$ and choose $0 < \hat{\tau} < \infty$ such that $M_{(m-X-\varepsilon)}(\hat{\tau}) < 1$. (Note that τ and $\hat{\tau}$ need not to be the minimizers.)

$$\begin{aligned}
\Pr[(X_1 - m - \varepsilon) + \cdots + (X_n - m - \varepsilon) \geq 0] &= \Pr\left[e^{\tau[(X_1-m-\varepsilon)+\cdots+(X_n-m-\varepsilon)]} \geq 1\right] \\
&\leq E^n[e^{\tau(X-m-\varepsilon)}] \quad \text{(by Markov's inequality)} \\
&= M^n_{(X-m-\varepsilon)}(\tau),
\end{aligned}$$

Strong law of large number revisited

and

$$\begin{aligned}
\Pr[(m - X_1 - \varepsilon) + \cdots + (m - X_n - \varepsilon) \geq 0] &= \Pr\left[e^{\tau[(m-X_1-\varepsilon)+\cdots+(m-X_n-\varepsilon)]} \geq 1\right] \\
&\leq E^n[e^{\tau(m-X-\varepsilon)}] \quad \text{(by Markov's inequality)} \\
&= M^n_{(m-X-\varepsilon)}(\tau).
\end{aligned}$$

□

Note that the validity of the upper bound only requires the assumption of "mean being negative."

Theorem 9.3 (Chernoff's Theorem) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables satisfying $E[X_n] < 0$ (and $E[X_n] > -\infty$), $P[X_n > 0] > 0$ (or $\inf_{t\in\Re} M_X(t) = M_X(\tau)$ for some finite τ) and all moments of X exists and can be determined by taking the derivative of its moment generating function. Then

$$\lim_{n\to\infty} -\frac{1}{n}\log\Pr\left[\frac{X_1+X_2+\cdots+X_n}{n} > 0\right] = -\inf_{t\in\Re}\log M_X(t) = -\inf_{t\in\Re} C_X(t) = I_X(0),$$

where $I_X(x) = \sup_{t\in\Re}[tx - C_X(t)]$.

Theorem (Cramér's Theorem) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables satisfying $M_X(t) < \infty$ for all t. Then

$$\begin{aligned}\inf_{x\geq a} I_X(x) &\leq \liminf_{n\to\infty} -\frac{1}{n}\log\Pr\left[\frac{X_1+X_2+\cdots+X_n}{n} > a\right] \\ &\leq \limsup_{n\to\infty} -\frac{1}{n}\log\Pr\left[\frac{X_1+X_2+\cdots+X_n}{n} > a\right] \leq \inf_{x>a} I_X(x)\end{aligned}$$

Law of the Iterated Logarithm

- The previous discussion introduces the rate of convergence for $\Pr[X_1 + X_2 + \cdots + X_n > 0]$ for i.i.d. $\{X_j\}_{j=1}^n$ and $-\infty < E[X_j] < 0$.
- We are thus confident that $\frac{1}{n}(X_1+\cdots+X_n)$ converges to its mean exponentially fast (in the sense that the probability of $\frac{1}{n}(X_1 + \cdots + X_n)$ deviating from its mean decreases to zero exponentially fast).
- The above result is obtained by *twisting* the original distribution with a specially chosen factor τ, resulting a larger $\Pr[Y^{(\tau)} > 0]$ with $E[Y^{(\tau)}] = 0$.

  Notably, if $Y = X_1 + X_2 + \cdots + X_n$, then $Y^{(\tau)} = X_1^{(\tau)} + X_2^{(\tau)} + \cdots + X_n^{(\tau)}$, and $\{X_i^{(\tau)}\}_{i=1}^n$ are also i.i.d.

  Claim: (i) $\{X_j\}_{j=1}^2$ i.i.d. (ii) Define a random variable $Y = X_1 + X_2$. (iii) Define new i.i.d. variables $\{Z_j\}_{j=1}^2$, where Z_j has distribution $P_{X^{(\tau)}}$. (iv) Define a new random variable $W = Z_1 + Z_2$.
  Then, $Y^{(\tau)}$ and W have exactly the same distribution. I.e., to compute $\Pr[Y^{(\tau)} \in \mathcal{A}]$ for any measurable set $\mathcal{A}$ can be implemented by $\Pr[X_1^{(\tau)} + X_2^{(\tau)} \in \mathcal{A}]$.

Law of the Iterated Logarithm

Proof:

$$\Pr[Y^{(\tau)} \le b] = \int_{-\infty}^{b} dP_{Y^{(\tau)}}(y) = \int_{-\infty}^{b} \frac{e^{\tau y}}{M_Y(\tau)} dP_Y(y) = \frac{1}{M_X^2(\tau)} \int_{-\infty}^{b} e^{\tau y} dP_Y(y)$$

$$\begin{aligned}
\Pr[W \le b] &= \int_{-\infty}^{\infty} \int_{-\infty}^{b-x_1} dP_{X^{(\tau)}}(x_2) dP_{X^{(\tau)}}(x_1) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{b-x_1} \frac{e^{\tau x_2}}{M_X(\tau)} dP_X(x_2) \frac{e^{\tau x_1}}{M_X(\tau)} dP_X(x_1) \\
&= \frac{1}{M_X^2(\tau)} \int_{-\infty}^{\infty} \int_{-\infty}^{b-x_1} e^{\tau(x_1+x_2)} dP_X(x_2) dP_X(x_1)
\end{aligned}$$

The equality of the above two terms can then be proved by noting that for a function $f(\cdot)$,

$$\int_{-\infty}^{b} f(y) dP_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{b-x_1} f(x_1 + x_2) dP_X(x_2) dP_X(x_1).$$

□

- We then provide a lower bound to $\Pr[Y^{(\tau)} > 0] = \Pr[X_1^{(\tau)} + X_2^{(\tau)} + \cdots + X_n^{(\tau)} > 0]$, where $E[X_j^{(\tau)}] = 0$ for $1 \le j \le n$. See Slide 9-29.

Law of the Iterated Logarithm

- **Question**: Can we say more about $X_1^{(\tau)} + X_2^{(\tau)} + \cdots + X_n^{(\tau)}$ with $\{X_j^{(\tau)}\}_{j=1}^n$ i.i.d. and $E[X_j^{(\tau)}] = 0$?

Law of the Iterated Logarithm

If $\{X_j\}_{j=0}^{\infty}$ are i.i.d. with zero mean and finite variance σ^2 (and implicitly all moments of X exist and can be determined by taking the derivatives of the moment generating function), then

1. $$\Pr\left[\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = 0\right] = 1 \quad \text{(Strong law of large numbers)}$$

2. $$\Pr\left[\limsup_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{2n\log\log(n)}} = 1\right] = 1 \quad \text{(Law of iterated logarithm)}$$

3. $$\lim_{n\to\infty} \Pr\left[\frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}} \le y\right] = \Phi(y) \quad \text{(Central limit theorem)}$$

where $\Phi(\cdot)$ is the cdf of the zero-mean unit-variance Gaussian distribution.

Implication The law of the iterated logarithm tells us that sum of all i.i.d. samples, normalized by $\sigma\sqrt{2n\log\log(n)}$, always approaches 1 infinitely often, regardless of the marginal distribution!

Law of the Iterated Logarithm

- $\limsup_{n\to\infty} a_n = a$ if, and only if, $(\forall\ \varepsilon > 0)(\exists\ N)(\forall\ n > N)\ a_n - a < \varepsilon$ and $(\forall\ \varepsilon > 0)(\forall\ N)(\exists\ n > N)\ a_n - a > -\varepsilon$.

> **Recall** $\lim_{n\to\infty} a_n = a$ if, and only if, $(\forall\ \varepsilon > 0)(\exists\ N)(\forall\ n > N)\ a_n - a < \varepsilon$ and $(\forall\ \varepsilon > 0)(\exists\ N)(\forall\ n > N)\ a_n - a > -\varepsilon$.

- Indeed, $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} = \lim_{n\to\infty} \sup_{k\geq n} a_k$.

Hence, to prove

$$\Pr\left[\limsup_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{2n\log\log(n)}} = 1\right] = 1$$

it suffices to prove that for every positive ε, (where ε is understood as "countable" in its range),

$$P\left(\left\{\underline{x} \in \Re^\infty : (\exists\ N)(\forall\ n > N)\frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log(n)}} - 1 < \varepsilon\right\}\right) = 1$$

and

$$P\left(\left\{\underline{x} \in \Re^\infty : (\forall\ N)(\exists\ n > N)\frac{x_1 + x_2 + \cdots + x_n}{\sigma\sqrt{2n\log\log(n)}} - 1 > -\varepsilon\right\}\right) = 1.$$

Law of the Iterated Logarithm

Equivalently,

$$P\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log(n)}}-1<\varepsilon\right\}\right)=1$$

and

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log(n)}}-1>-\varepsilon\right\}\right)=1.$$

Or equivalently, by De Morgan's law,

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log(n)}}-1\geq\varepsilon\right\}\right)=0$$

and

$$P\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\frac{x_1+x_2+\cdots+x_n}{\sigma\sqrt{2n\log\log(n)}}-1>-\varepsilon\right\}\right)=1.$$

Law of the Iterated Logarithm

Or equivalently (as written in Billingsley's book),

$$\Pr\left[X_1 + X_2 + \cdots + X_n \geq (1+\varepsilon)\sigma\sqrt{2n\log\log(n)} \text{ i.o.}\right] = 0$$

and

$$\Pr\left[X_1 + X_2 + \cdots + X_n > (1-\varepsilon)\sigma\sqrt{2n\log\log(n)} \text{ i.o.}\right] = 1.$$

In order to reduce the confusion, I personally prefer to distinguish between the "event" and the "set". That is why I put "Pr[event]" and use "P(set)", where the event is defined through some random variables, and the probability of a set is measured by probability measure P. Billingsley's book sometimes mixes the two together, which may be confused for beginners.

Assumption Without loss of generality, we assume that the variance σ^2 is unity.

Proof of Law of the Iterated Logarithm

In order to prove the previous two equalities, we need two preliminary theorems.

Theorem 9.4 Let $S_n = X_1 + X_2 + \cdots + X_n$, where $\{X_j\}_{j=1}^{\infty}$ are i.i.d. with zero mean and unit variance (and implicitly all moments of X_j exist and can be determined by taking the derivative of the moment generating function). Suppose that the positive constant sequence $a_1, a_2, \ldots, a_n, \ldots$ satisfies

$$a_n \to \infty \quad \text{and} \quad \frac{a_n}{\sqrt{n}} \to 0.$$

Then there exists a sequence $\zeta_1, \zeta_2, \ldots$ with $\zeta_n \to 0$ such that

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

Recall $S_n/\sqrt{n}$ converges in distribution to zero-mean unit-variance Gaussian. So

$$\begin{aligned} P\left[S_n \geq a\sqrt{n}\right] &\to \int_a^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds \\ &= \frac{1}{\sqrt{2\pi}a} e^{-a^2/2}\left(1 - \frac{1}{a^2} + \frac{1\cdot 3}{a^4} - \frac{1\cdot 3\cdot 5}{a^6} + \cdots\right), \end{aligned}$$

where the last equality holds for $a > 0$. In this theorem, we divide S_n by a little larger quantity than $\sqrt{n}$.

Proof of Law of the Iterated Logarithm

Proof: Observe that:

$$\begin{aligned}\Pr[S_n \geq a_n\sqrt{n}] &= \Pr[X_1 + X_2 + \cdots + X_n \geq a_n\sqrt{n}] \\ &= \Pr[(X_1 - a_n/\sqrt{n}) + (X_2 - a_n/\sqrt{n}) + \cdots + (X_n - a_n/\sqrt{n}) \geq 0].\end{aligned}$$

Then by letting $Y = (X_1 - a_n/\sqrt{n}) + (X_2 - a_n/\sqrt{n}) + \cdots + (X_n - a_n/\sqrt{n})$ and noting that $\{X_j - a_n/\sqrt{n}\}_{j=1}^n$ are i.i.d., we have $-\infty < E[Y] = -a_n\sqrt{n} < 0$ and for all sufficiently large n,

$$\begin{aligned}&\Pr[Y > 0] \\ &\geq \Pr[(X_1 - a_n/\sqrt{n} > 0) \text{ and } (X_2 - a_n/\sqrt{n}) > 0 \text{ and } \cdots \text{ and } (X_n - a_n/\sqrt{n}) > 0] \\ &= \left(\Pr[X_1 - a_n/\sqrt{n} > 0]\right)^n > 0 \text{ (The strict positivity is proved below.)}\end{aligned}$$

We claim that $(\exists\ \varepsilon > 0) \Pr[X_1 > \varepsilon] > 0$ because if it were not true, i.e., if $(\forall\ \varepsilon > 0) \Pr[X_1 > \varepsilon] = 0$, then $\Pr[X_1 \leq 0] = 1$, which together with $E[X_1] = 0$ implies that $\Pr[X_1 = 0] = 1$, which violates $E[X_1^2] = 1$. Accordingly, as $a_n/\sqrt{n} \to 0$, there exists N such that for $n > N$, $a_n/\sqrt{n} < \varepsilon$, and hence, for $n > N$, $\Pr[X_1 > a_n/\sqrt{n}] \geq \Pr[X_1 > \varepsilon] > 0$. □

So we can apply the previously derived upper and lower bounds to $\Pr[Y \geq 0]$.

Proof of Law of the Iterated Logarithm

Definition (o-notation) Suppose that $g(x) \neq 0$ for all $x \neq a$ in some open interval containing a. Then $f(x)$ is little-oh of $g(x)$ or $f(x)$ is of smaller order than $g(x)$, denoted by

$$f(x) = o(g(x)) \text{ as } x \to a,$$

if

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0.$$

Definition (O-notation) Suppose that $g(x) \neq 0$ for all $x \neq a$ in some open interval containing a. $f(x)$ is big-oh of $g(x)$ or $f(x)$ is of the same order as $g(x)$, denoted by

$$f(x) = O(g(x)) \text{ as } x \to a,$$

if

$$\lim_{\varepsilon \to 0} \sup_{\{x \in \Re : |x-a| < \varepsilon\}} \left| \frac{f(x)}{g(x)} \right| < \infty.$$

$$M_Y(\tau_n) \geq \Pr[Y \geq 0] \geq M_Y(\tau_n) \cdot \frac{E^2[(Y^{(\tau_n)})^2]}{4E[(Y^{(\tau_n)})^4]} \exp\left\{-\frac{4\tau_n E[(Y^{(\tau_n)})^4]}{E^{3/2}[(Y^{(\tau_n)})^2]}\right\}.$$

- *Upper bound*: $M_Y(t) = \prod_{i=1}^{n} M_{(X_i - a_n/\sqrt{n})}(t) = M^n_{(X-a_n/\sqrt{n})}(t)$

$\Rightarrow M'_Y(t) = nM'_{(X-a_n/\sqrt{n})}(t) M^{n-1}_{(X-a_n/\sqrt{n})}(t)$

$\Rightarrow \tau_n$ satisfies $M'_{(X-a_n/\sqrt{n})}(\tau_n) = 0$.

Equivalently, τ_n satisfies $C'_{(X-a_n/\sqrt{n})}(\tau_n) = C'_X(\tau_n) - a_n/\sqrt{n} = 0$.

The zero-mean assumption implies that $C'_X(0) = 0$. Also, the strictly positive second moment indicates the strict convexity of $C_X(t)$, which in turn implies the strict increasingness of $C'_X(t)$ in t. As a result, $a_n/\sqrt{n} \to 0$ as $n \to \infty$ gives that $\tau_n \to 0$ as $n \to \infty$, because $C'_X(\tau_n) = a_n/\sqrt{n}$.

By Taylor expansion,

$$C_X(t) = C_X(0) + C'_X(0)t + \frac{C''_X(0)}{2!}t^2 + O(t^3) = \frac{1}{2}t^2 + O(t^3) \text{ as } t \to 0,$$

from which we obtain (by taking the derivative)

$$C'_X(\tau_n) = \tau_n + O(\tau_n^2) = a_n/\sqrt{n} \text{ as } \tau_n \to 0 \text{ (or as } n \to \infty),$$

Proof of Law of the Iterated Logarithm

which implies $\tau_n = a_n/\sqrt{n} + O(a_n^2/n) \to 0$ as $n \to \infty$; hence, τ_n is eventually **finite**.

$\limsup_{n\to\infty} \frac{|\tau_n - a_n/\sqrt{n}|}{\tau_n^2} < C$ implies $(\exists\ N)(\forall\ n > N)|\tau_n - a_n/\sqrt{n}| < C\tau_n^2$, or equivalently,

$$\tau_n - C\tau_n^2 < a_n/\sqrt{n} < \tau_n + C\tau_n^2.$$

Let N_1 be the smallest n satisfying $1 - C\tau_n > 1/2$ for all $n > N_1$. Then for $n > \max\{N, N_1\}$, $a_n^2/n > \tau_n^2(1 - C\tau_n)^2$, which concludes that

$$(\forall\ n > \max\{N, N_1\})\frac{|\tau_n - a_n/\sqrt{n}|}{a_n^2/n} < \frac{C\tau_n^2}{\tau_n^2(1 - C\tau_n)^2} = \frac{C}{(1 - C\tau_n)^2} < 4C.$$

(It is the reason why the theorem assumes $a_n/\sqrt{n} \to 0$; the theorem does require $\tau_n \to 0$.)

Notably, τ_n is **dependent** on n.

Proof of Law of the Iterated Logarithm

$$\log \rho_n = \log M_{(X-a_n/\sqrt{n})}(\tau_n) = C_{(X-a_n/\sqrt{n})}(\tau_n) = C_X(\tau_n) - a_n\tau_n/\sqrt{n}$$
$$= \left[\frac{1}{2}\tau_n^2 + O(\tau_n^3)\right] - [\tau_n + O(\tau_n^2)]\tau_n = -\frac{1}{2}\tau_n^2 + O(\tau_n^3) \text{ as } n \to \infty.$$

We then obtain:

$$\rho_n^n \geq \Pr[Y \geq 0]$$

$$M_Y(\tau_n) \geq \Pr[Y \geq 0] \geq M_Y(\tau_n) \cdot \frac{E^2[(Y^{(\tau_n)})^2]}{4E[(Y^{(\tau_n)})^4]} \exp\left\{ -\frac{4\tau_n E[(Y^{(\tau_n)})^4]}{E^{3/2}[(Y^{(\tau_n)})^2]} \right\}.$$

- *Lower bound*:
  - $M_Y(t) = M^n_{(X-a_n/\sqrt{n})}(t)$
    $\Rightarrow \underline{M_Y(\tau_n)} = M^n_{(X-a_n/\sqrt{n})}(\tau_n) = \underline{\rho_n^n}$
  - $C_Y(t) = nC_{(X-a_n/\sqrt{n})}(t) = n[C_X(t) - a_n t/\sqrt{n}]$
  - $E[(Y^{(\tau_n)})^2] = C''_{Y^{(\tau_n)}}(0) = C''_Y(\tau_n) = nC''_X(\tau_n) = n\lambda_{2,n}$,
    where $\lambda_{2,n} = C''_X(\tau_n)$.

Proof of Law of the Iterated Logarithm

– $$E[(Y^{(\tau_n)})^4] = C^{(4)}_{Y^{(\tau_n)}}(0) + 3E^2[(Y^{(\tau_n)})^2]$$
$$= C^{(4)}_{Y}(\tau_n) + 3n^2\lambda^2_{2,n}$$
$$= nC^{(4)}_{X}(\tau_n) + 3n^2\lambda^2_{2,n}$$
$$= n\lambda_{4,n} + 3n^2\lambda^2_{2,n},$$

where $\lambda_{4,n} = C^{(4)}_X(\tau_n)$.

– Then

$$\frac{E^2[(Y^{(\tau_n)})^2]}{4E[(Y^{(\tau_n)})^4]} = \frac{E^2[(Y^{(\tau_n)})^2]/n^2}{4E[(Y^{(\tau_n)})^4]/n^2} = \frac{\lambda^2_{2,n}}{4\left(3\lambda^2_{2,n} + \lambda_{4,n}/n\right)} = \frac{1}{12 + 4\lambda_{4,n}/(\lambda^2_{2,n}n)}$$

and

$$\frac{4\tau_n E[(Y^{(\tau_n)})^4]}{E^{3/2}[(Y^{(\tau_n)})^2]} = \frac{4\tau_n[n\lambda_{4,n} + 3n^2\lambda^2_{2,n}]}{n^{3/2}\lambda^{3/2}_{2,n}} = 4\tau_n\left[3 + \lambda_{4,n}/(\lambda^2_{2,n}n)\right]\sqrt{\lambda_{2,n}n}$$

Accordingly,

$$\Pr[Y \geq 0] \;\geq\; \rho_n^n \cdot \frac{1}{12 + 4\lambda_{4,n}/(\lambda^2_{2,n}n)} \exp\left\{-4\tau_n\left[3 + \lambda_{4,n}/(\lambda^2_{2,n}n)\right]\sqrt{\lambda_{2,n}n}\right\}$$

Proof of Law of the Iterated Logarithm

This concludes to:

$$\rho_n^n \geq \Pr[Y \geq 0] \geq \rho_n^n \cdot \frac{1}{12 + 4\lambda_{4,n}/(\lambda_{2,n}^2 n)} \exp\left\{-4\tau_n \left[3 + \lambda_{4,n}/(\lambda_{2,n}^2 n)\right] \sqrt{\lambda_{2,n} n}\right\},$$

where (as $n \to \infty$ or $t \to 0$ for the O-notation)

$$\begin{cases} C_X(t) = \frac{1}{2}t^2 + O(t^3) \quad \left(= \frac{1}{2}t^2 + \frac{E[X^3]}{3!}t^3 + O(t^4) \text{ for the computation of } C_X^{(4)}(t)\right) \\ \tau_n = \frac{a_n}{\sqrt{n}} + O\left(\frac{a_n^2}{n}\right) \qquad \left(\text{or } \frac{a_n}{\sqrt{n}} = \tau_n + O(\tau_n^2)\right) \\ \log \rho_n = -\frac{1}{2}\tau_n^2 + O(\tau_n^3) = -\frac{a_n^2}{2n} + O\left(\frac{a_n^3}{n^{3/2}}\right) \\ \lambda_{2,n} = C_X''(\tau_n) = 1 + O(\tau_n) = 1 + O\left(\frac{a_n}{\sqrt{n}}\right) \\ \lambda_{4,n} = C_X^{(4)}(\tau_n) = O(1) \end{cases}$$

Hence, $\dfrac{\lambda_{4,n}}{\lambda_{2,n}^2} = \dfrac{O(1)}{\left[1 + O\left(\frac{a_n}{\sqrt{n}}\right)\right]^2} = O(1)$ as $n \to \infty$.

As a result, by letting $\Pr[Y \geq 0] = e^{-a_n^2(1+\zeta_n)/2}$, we obtain (as $n \to \infty$ for the O-notation)

$$n \log \rho_n \geq -\frac{1}{2}a_n^2(1+\zeta_n) \geq n \log \rho_n - \log\left(12 + 4\frac{\lambda_{4,n}}{\lambda_{2,n}^2 n}\right) - 4\tau_n\left(3 + \frac{\lambda_{4,n}}{\lambda_{2,n}^2 n}\right)\sqrt{\lambda_{2,n} n}$$

$$\Leftrightarrow \quad -\frac{a_n^2}{2} + O\left(\frac{a_n^3}{n^{1/2}}\right) \geq -\frac{1}{2}a_n^2(1+\zeta_n) \geq -\frac{a_n^2}{2} + O\left(\frac{a_n^3}{n^{1/2}}\right)$$

$$-\log\left(12 + O\left(\frac{1}{n}\right)\right) - 4\left(\frac{a_n}{\sqrt{n}} + O\left(\frac{a_n^2}{n}\right)\right)\left[3 + O\left(\frac{1}{n}\right)\right]\sqrt{\left(1 + O\left(\frac{a_n}{\sqrt{n}}\right)\right)n}$$

$$\Leftrightarrow \quad O\left(\frac{a_n}{\sqrt{n}}\right) \leq \zeta_n \leq O\left(\frac{a_n}{\sqrt{n}}\right)$$

$$+\frac{2}{a_n^2}\log\left(12 + O\left(\frac{1}{n}\right)\right) + 8\left(\frac{1}{a_n} + O\left(\frac{1}{\sqrt{n}}\right)\right)\left[3 + O\left(\frac{1}{n}\right)\right]\sqrt{1 + O\left(\frac{a_n}{\sqrt{n}}\right)}$$

Thus, if $a_n \to \infty$ and $a_n/\sqrt{n} \to 0$, then $\zeta_n \to 0$. □

Theorem 9.6 Let $S_n = X_1 + \cdots + X_n$, where $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with mean 0 and variance 1 (and all moments of X_i exist, which can be determined by taking the derivatives of its moment generating function). Also, let $\Pr[S_0 = 0] = 1$. Then for $\alpha \geq \sqrt{2}$,

$$\Pr\left[\frac{\max\{S_0, S_1, \ldots, S_n\}}{\sqrt{n}} \geq \alpha\right] \leq 2 \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right].$$

Proof: Denote $M_n = \max\{S_0, S_1, \ldots, S_n\}$. Then M_n is non-negative and non-decreasing in n.

$$\begin{aligned}
\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-1}}{\sqrt{n}} < \alpha\right)\right. \\
&\quad \vee \left(\frac{M_n}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-1}}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-2}}{\sqrt{n}} < \alpha\right) \vee \cdots \\
&\quad \left. \vee \left(\frac{M_n}{\sqrt{n}} \geq \alpha \wedge \frac{M_{n-1}}{\sqrt{n}} \geq \alpha \wedge \cdots \wedge \frac{M_1}{\sqrt{n}} \geq \alpha \wedge \frac{M_0}{\sqrt{n}} < \alpha\right)\right] \\
&= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha > \frac{M_{n-1}}{\sqrt{n}}\right) \vee \left(\frac{M_{n-1}}{\sqrt{n}} \geq \alpha > \frac{M_{n-2}}{\sqrt{n}}\right) \vee \cdots \vee \left(\frac{M_1}{\sqrt{n}} \geq \alpha > \frac{M_0}{\sqrt{n}}\right)\right] \\
&= \sum_{j=1}^{n} \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right].
\end{aligned}$$

Proof of Law of the Iterated Logarithm

Using this result, we can further derive:

$$\begin{aligned}
\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right)\right] \\
&\qquad + \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&= \Pr\left[\left(\frac{M_n}{\sqrt{n}} \geq \alpha\right) \wedge \left(\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right)\right] \\
&\qquad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] \\
&\qquad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right].
\end{aligned}$$

Proof of Law of the Iterated Logarithm

Since $\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right)$ implies $\frac{S_j}{\sqrt{n}} \geq \alpha$, we have:

$$\Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right]$$

$$\leq \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \frac{S_j}{\sqrt{n}} - \sqrt{2}\right)\right]$$

$$= \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_j - S_n}{\sqrt{n}} > \sqrt{2}\right)\right]$$

$$= \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \Pr\left[\frac{S_j - S_n}{\sqrt{n}} > \sqrt{2}\right]$$

(M_j and M_{j-1} only depend on $X_1, \ldots, X_j$; $S_j - S_n$ only depends on $X_{j+1}, \ldots, X_n$. So the above two events are independent.)

$$\leq \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \Pr\left[\frac{|S_j - S_n|}{\sqrt{n}} > \sqrt{2}\right]$$

$$\leq \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \frac{\mathrm{Var}[S_n - S_j]}{2n} \quad \text{(by Chebyshev's ineq.)}$$

$$= \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \frac{(n-j)}{2n} \leq \frac{1}{2} \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right].$$

Proof of Law of the Iterated Logarithm

Consequently,

$$
\begin{aligned}
\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right] &\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] \\
&\quad + \sum_{j=1}^{n} \Pr\left[\left(\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right) \wedge \left(\frac{S_n}{\sqrt{n}} < \alpha - \sqrt{2}\right)\right] \\
&\leq \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \frac{1}{2}\sum_{j=1}^{n} \Pr\left[\frac{M_j}{\sqrt{n}} \geq \alpha > \frac{M_{j-1}}{\sqrt{n}}\right] \\
&= \Pr\left[\frac{S_n}{\sqrt{n}} \geq \alpha - \sqrt{2}\right] + \frac{1}{2}\Pr\left[\frac{M_n}{\sqrt{n}} \geq \alpha\right].
\end{aligned}
$$

□

Proof of Law of the Iterated Logarithm

Theorem 9.5 (Law of the Iterated Logarithm) Let $S_n = X_1 + \cdots + X_n$, where $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with mean 0 and variance 1 (and all moments of X_i exist, which can be determined by taking the derivatives of its moment generating function). Then

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

Proof: We shall prove the two equations in Slide 9-53.

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

I have modified the proof in Billingsley's book so that the proof is more "straightforward" for engineering major students.

Choose a positive θ such that $1 < \theta^3 < 1+\varepsilon$.

Let $n_k \triangleq \lfloor \theta^k \rfloor > \theta^k - 1$, and

for $k \geq \log_\theta(e^e + 1)$ (i.e., $\theta^k - 1 \geq e^e$, which implies $\log\log(n_k) \geq 1$),

let $x_k \triangleq \theta\sqrt{2\log\log(n_k)}\ (\geq \theta\sqrt{2}) \geq \sqrt{2}$.

Proof of Law of the Iterated Logarithm

Then Theorems 9.4 and 9.6 give that for $k \geq \log_\theta(e^e + 1)$ and for some $\zeta_k \to 0$,

$$\begin{aligned}
\Pr\left[\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right] &\leq 2\Pr\left[\frac{S_{n_k}}{\sqrt{n_k}} \geq x_k - \sqrt{2}\right] \quad \text{(Theorem 9.6)} \\
&= 2\exp\left\{-\frac{1}{2}(x_k - \sqrt{2})^2(1+\zeta_k)\right\} \quad \text{(Theorem 9.4)} \\
&= 2\exp\left\{-\theta^2\log(k)\frac{(x_k - \sqrt{2})^2(1+\zeta_k)}{2\,\theta^2\log(k)}\right\}.
\end{aligned}$$

Since

$$\begin{aligned}
\lim_{k\to\infty}\frac{(x_k - \sqrt{2})^2(1+\zeta_k)}{2\theta^2\log(k)} &= \lim_{k\to\infty}\frac{(\theta\sqrt{2\log\log(\theta^k)} - \sqrt{2})^2}{2\theta^2\log(k)} \quad \text{(Because } \zeta_k \downarrow 0\text{)} \\
&= \lim_{k\to\infty}\frac{(\theta\sqrt{\log\log(\theta^k)} - 1)^2}{\theta^2\log(k)} \\
&= \lim_{k\to\infty}\frac{(\theta\sqrt{\log(k) + \log\log(\theta)} - 1)^2}{\theta^2\log(k)} \\
&= 1,
\end{aligned}$$

there exists K such that for $k \geq K$,

$$\frac{(x_k - \sqrt{2})^2(1+\zeta_k)}{2\theta^2\log(k)} \geq \frac{1}{\theta}.$$

Proof of Law of the Iterated Logarithm

Accordingly, for $k \geq K_0 \triangleq \max\{K, \log_\theta(e^e + 1)\}$,

$$\Pr\left[\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right] \leq 2\exp\left\{-\theta\log(k)\right\} = \frac{2}{k^\theta}.$$

Now for $n \geq e^{e^\theta}$ fixed, there exists k such that $n_{k-1} < n \leq n_k$, and

$$\sqrt{2n\log\log(n)} \geq \sqrt{2(n_{k-1}+1)\log\log(n_{k-1}+1)} > \sqrt{2\left(\frac{n_k}{\theta}\right)\log\log\left(\frac{n_k}{\theta}\right)},$$

The last strict inequality follows from:

$$n_{k-1} = \lfloor\theta^{k-1}\rfloor > \theta^{k-1} - 1 = \frac{\theta^k}{\theta} - 1 \geq \frac{\lfloor\theta^k\rfloor}{\theta} - 1 = \frac{n_k}{\theta} - 1.$$

which implies that (This is the only step requiring $\theta^3 < (1+\varepsilon)$)

$$\begin{aligned}(1+\varepsilon)\sqrt{2n\log\log(n)} &> \theta^3\sqrt{2\left(\frac{n_k}{\theta}\right)\log\log\left(\frac{n_k}{\theta}\right)}\\ &= \theta^2\sqrt{2n_k\theta\log\log\left(\frac{n_k}{\theta}\right)}\\ &\geq \theta\sqrt{2n_k\log\log\left(n_k\right)} \quad (\text{since } \theta^2 > \theta)\end{aligned}$$

Proof of Law of the Iterated Logarithm

Claim: Given $\theta > 1$, $f_\theta(x) = \log\log\left(\frac{x}{\theta}\right) - \frac{1}{\theta}\log\log(x) \geq 0$ for $x \geq e^{e^\theta}$.

Proof: The claim can be validated by $f'_\theta(x) = \dfrac{(\theta-1)\log(x) + \log(\theta)}{x\theta\log(x)[\log(x) - \log(\theta)]} > 0$ for $x \geq e^{e^\theta}$ and $f_\theta(e^{e^\theta}) = \log[e^\theta - \log(\theta)] - 1 > 0$ for any $\theta > 1$. $\square$

Accordingly, $\left[S_n \geq (1+\varepsilon)\sqrt{2n\log\log(n)}\right]$ implies that

$$M_{n_k} \quad \left(\geq S_n \geq (1+\varepsilon)\sqrt{2n\log\log(n)}\right) \quad \geq \theta\sqrt{2n_k\log\log\left(n_k\right)},$$

where k is the unique integer satisfying $\lfloor\theta^{k-1}\rfloor < n \leq \lfloor\theta^k\rfloor$ (namely, $k = \lceil\log_\theta(n)\rceil$). As a consequence,

$$\Pr\left[S_n \geq (1+\varepsilon)\sqrt{2n\log\log(n)} \text{ i.o. in } n\right] \leq \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k\log\log\left(n_k\right)} \text{ i.o. in } k\right].$$

Theorem 4.3 (First Borel-Cantelli Lemma) If $\sum_{n=1}^{\infty} P(A_n)$ converges (i.e., $\sum_{n=1}^{\infty} P(A_n) < \infty$), then $P\left(\limsup_{n\to\infty} A_n\right) = P(A_n \text{ i.o.}) = 0$.

Proof of Law of the Iterated Logarithm

By the first Borel-Cantelli lemma,

$$\begin{aligned}\sum_{k=1}^{\infty} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] &= \sum_{k=1}^{K_0} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] \\ &\quad + \sum_{k=K_0+1}^{\infty} \Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)}\right] \\ &\leq K_0 + \sum_{k=K_0+1}^{\infty} \frac{2}{k^\theta} \\ &\leq K_0 + \int_{K_0}^{\infty} \frac{2}{x^\theta} dx \\ &= K_0 + \frac{2}{(\theta-1)} \frac{1}{K_0^{\theta-1}} \\ &< \infty,\end{aligned}$$

we obtain (The below equality holds without the condition that $\theta^3 < (1+\varepsilon)$.):

$$\Pr\left[M_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)} \text{ i.o. in } k\right] = 0.$$

This completes the proof of the first part.

Proof of Law of the Iterated Logarithm

2. $\Pr\left[S_n > (1-\varepsilon)\sqrt{2n\log\log(n)} \text{ i.o.}\right] = 1.$

Choose ξ satisfying $\xi > \max\{1, 9/\varepsilon^2\}$. Take $n_k \triangleq \lfloor \xi^k \rfloor$.

For any k, let $m_k = n_k - n_{k-1}$ and $a_k = \dfrac{\left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{\sqrt{m_k}}$.

Then $a_k \to \infty$ and $\dfrac{a_k}{\sqrt{m_k}} \to 0$ as $k \to \infty$.

- $a_k \geq \dfrac{\left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{\sqrt{n_k}} = \left(1-\dfrac{1}{\xi}\right)\sqrt{2\log\log(n_k)} \to \infty.$
-

$$\begin{aligned}\lim_{k\to\infty}\frac{a_k}{\sqrt{m_k}} &= \lim_{k\to\infty}\frac{\left(1-\frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{m_k}\\ &= \lim_{k\to\infty}\frac{\left(1-\frac{1}{\xi}\right)\sqrt{2\xi^k\log\log(\xi^k)}}{\xi^k-\xi^{k-1}} = \lim_{k\to\infty}\sqrt{\frac{2\log\log(\xi^k)}{\xi^k}} = 0\end{aligned}$$

Theorem 9.4 then implies:

$$
\begin{aligned}
&\Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&= \Pr\left[X_{n_{k-1}+1} + \cdots + X_{n_k} \geq a_k\sqrt{m_k}\right] \\
&= \exp\left\{-\frac{1}{2}a_k^2(1+\zeta_k)\right\} \text{ for some } \zeta_k \to 0 \\
&= \exp\left\{-\frac{1}{2}\left(\frac{\left(1 - \frac{1}{\xi}\right)\sqrt{2n_k\log\log(n_k)}}{\sqrt{m_k}}\right)^2(1+\zeta_k)\right\} \\
&= \exp\left\{-\frac{(\xi-1)^2[n_k\log\log(n_k)](1+\zeta_k)}{\xi^2(n_k - n_{k-1})}\right\} \\
&\geq \exp\left\{-\frac{(\xi-1)^2[\xi^k\log\log(\xi^k)](1+\zeta_k)}{\xi^2([\xi^k-1]-\xi^{k-1})}\right\} \quad (\text{since } \xi^k - 1 < n_k \leq \xi^k \text{ and } n_{k-1} \leq \xi^{k-1}) \\
&= \exp\left\{-\frac{(\xi-1)^2[\xi^k\log\log(\xi^k)](1+\zeta_k)}{\xi^2(\xi^k-\xi^{k-1})\left(1-\frac{1}{\xi^k-\xi^{k-1}}\right)}\right\} = \exp\left\{-\frac{(\xi-1)\log\log(\xi^k)(1+\zeta_k)}{\xi\left(1-\frac{1}{\xi^k-\xi^{k-1}}\right)}\right\}
\end{aligned}
$$

Proof of Law of the Iterated Logarithm

As $\zeta_k \to 0$, there exists K_1 such that $\zeta_k < \dfrac{1}{2\xi - 1}$ for all $k \geq K_1$.

Also, $k \geq \dfrac{\log(\xi) + \log(2\xi - 1) - \log(\xi - 1)}{\log(\xi)}$ if, and only if, $\dfrac{1}{\xi^k - \xi^{k-1}} \leq \dfrac{1}{2\xi - 1}$.

Hence, for $k \geq K_2 \triangleq \max\left\{K_1, \dfrac{\log(\xi) + \log(2\xi - 1) - \log(\xi - 1)}{\log(\xi)}\right\}$,

$$
\begin{aligned}
\Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] &\geq \exp\left\{-\frac{(\xi - 1)\log\log(\xi^k)(1 + \zeta_k)}{\xi\left(1 - \dfrac{1}{\xi^k - \xi^{k-1}}\right)}\right\} \\
&\geq \exp\left\{-\frac{(\xi - 1)\log\log(\xi^k)\left(1 + \dfrac{1}{2\xi - 1}\right)}{\xi\left(1 - \dfrac{1}{2\xi - 1}\right)}\right\} \\
&= \exp\left\{-\log\log(\xi^k)\right\} = \frac{1}{k\log(\xi)}.
\end{aligned}
$$

Proof of Law of the Iterated Logarithm

Theorem 4.4 (Second Borel-Cantelli Lemma) If $\{A_n\}_{n=1}^{\infty}$ forms an independent sequence of events for a probability measure P, and $\sum_{n=1}^{\infty} P(A_n)$ diverges, then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Since $\left\{\left[S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right]\right\}_{k=1}^{\infty}$ are independent events, and

$$\begin{aligned}
&\sum_{k=1}^{\infty} \Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&\geq \sum_{k=K_2}^{\infty} \Pr\left[S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}\right] \\
&\geq \sum_{k=K_2}^{\infty} \frac{1}{k\log(\xi)} = \infty,
\end{aligned}$$

it follows from the second Borel-Cantelli lemma that with probability 1, $S_{n_k} - S_{n_{k-1}} \geq \left(1 - \frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)}$ infinitely often in k.

Proof of Law of the Iterated Logarithm

Now we can let $\bar{X}_n = -X_n$, and let $\bar{M}_n$ and $\bar{S}_n$ be respectively the counterparts of M_n and S_n for $\{\bar{X}_n\}$ (and $n_k = \lfloor \xi^k \rfloor$), and apply the proof in the first part with $\theta = \sqrt{2}$ to show that (cf. Slide 9-72)(This holds without $\theta^3 < (1+\epsilon)$):

$$\Pr\left[\bar{M}_{n_k} \geq \theta\sqrt{2n_k \log\log(n_k)} \text{ i.o. in } k\right] = 0.$$

Hence, it is with probability 1 that $-S_{n_{k-1}} = \bar{S}_{n_{k-1}} \leq \bar{M}_{n_{k-1}} < \theta\sqrt{2n_{k-1}\log\log(n_{k-1})}$ for all but finitely many k.

Observe that $\theta\sqrt{2n_{k-1}\log\log(n_{k-1})} \leq \dfrac{2}{\sqrt{\xi}}\sqrt{2n_k \log\log(n_k)}$.

The validity of the above inequality follows:
Apply $n_{k-1}\left(\triangleq \lfloor \xi^{k-1} \rfloor \leq \xi^{k-1} = \dfrac{\xi^k}{\xi} \leq \dfrac{n_k+1}{\xi}\right) \leq \dfrac{2n_k}{\xi}$ for n_{k-1} outside $\log\log(\cdot)$.
Apply $n_{k-1} \leq n_k$ for n_{k-1} inside $\log\log(\cdot)$.

As a result, it is with probability 1 that $-S_{n_{k-1}} \leq \dfrac{2}{\sqrt{\xi}}\sqrt{2n_k\log\log(n_k)}$ for all but finitely many k.

To summarize, it is with probability 1 that for infinitely many k,

$$
\begin{aligned}
S_{n_k} &\geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)} + S_{n_{k-1}} \\
&\geq \left(1-\frac{1}{\xi}\right)\sqrt{2n_k \log\log(n_k)} - \frac{2}{\sqrt{\xi}}\sqrt{2n_k \log\log\left(n_k\right)} \\
&= \left(1-\frac{1}{\xi}-\frac{2}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&\geq \left(1-\frac{1}{\sqrt{\xi}}-\frac{2}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&= \left(1-\frac{3}{\sqrt{\xi}}\right)\sqrt{2n_k \log\log(n_k)} \\
&\geq (1-\varepsilon)\sqrt{2n_k \log\log(n_k)}.
\end{aligned}
$$

□

Final note

- The original statement of the law of the iterated logarithm was due to A. Y. Khinchin in 1924.
- Another statement of the law of the iterated logarithm was given by A. N. Kolmogorov in 1929.