Solution for Note 3 of 275A Probability Theory

shuyu qin

August 2023

1 Selected Exercises in Note 3

Given $\lim_N \mathbf{P}(\bigvee_{n\geq N}(|X_n - X| > \varepsilon)) = 0$ for every $\varepsilon > 0$, continuity from above implies that

$\mathbf{P}(\bigwedge_{N\geq 1}\bigvee_{n\geq N}(|X_n - X| > \varepsilon)) = \mathbf{P}(\limsup_n |X_n - X| > \varepsilon) = 0,$

hence by monotonicity, $\forall\varepsilon > 0$,

$\mathbf{P}(\liminf_n |X_n - X| > \varepsilon) = \mathbf{P}(\limsup_n |X_n - X| > \varepsilon) = 0,$

which by monotonicity again implies that $\mathbf{P}(\limsup_n X_n = \liminf_n X_n = X) = 1$. That is, $X_n \to X$ almost surely.

Conversely, if $X_n \to X$ almost surely, then $\mathbf{P}(\lim_n X_n \neq X) = 0$, so in particular $\forall\varepsilon > 0, \mathbf{P}(\limsup_n |X_n - X| > \varepsilon) = 0$, indicating

$\lim_N \mathbf{P}(\bigvee_{n\geq N} |X_n - X| > \varepsilon) = 0,\ \forall\varepsilon > 0.$

Exercise 2

Proof. We fix an $\varepsilon > 0$ throughout.

(1). By monotonicity, $\mathbf{P}(|X_n - X| > \varepsilon) \leq \mathbf{P}(\bigvee_{n\geq N}(|X_n - X| > \varepsilon))$ for $n \geq N$, sending $N \to \infty$ then gives the result. A counterexample is the typewriter sequence, which converges in measure (in probability) and in L^1 norm (in mean), but not pointwise almost everywhere (almost surely), see Example 1.5.5 of "M".

(2). See also Exercise 1.5.5 (Fast L^1 convergence) of "M".

Let $E_n := |X_n - X| > \varepsilon$ for any n, and $F_N := \bigvee_{n\geq N} E_n$. Suppose for contradiction that $\lim_{N\to\infty} \mathbf{P}(F_N) \neq 0$, then without loss of generality we can assume the sequence $\mathbf{P}(F_N)$ is bounded away from 0. That is, $\exists c > 0$ such that $\mathbf{P}(F_N) \geq c$ for all N. By construction, F_N is decreasing and so by the monotone convergence theorem we have $\lim_N \mathbf{P}(F_N) = d$ for some $d > 0$. By continuity from above, $\mathbf{P}(\bigwedge_{N=1}^{\infty} F_N) = d > 0$, but this contradicts the Borel-Cantelli lemma and so $X_n \to X$ almost surely.

A counterexample is the escape to horizontal infinity given in Example 1.5.2 of "M".

(3). See Exercise 1.5.6 of "M".

Let $\varepsilon' > 0$ be arbitrary, if $X_n \to X$ in probability, we can find an increasing sequence $(n_j)_j$ such that $\mathbf{P}(|X_{n_j} - X| > \varepsilon) < \varepsilon'/2^j$ for every j, by countable subadditivity:

$\mathbf{P}(\bigvee_{j \geq N} (|X_{n_j} - X| > \varepsilon)) \leq \sum_j \mathbf{P}(|X_{n_j} - X| > \varepsilon) < \varepsilon'$. That is, the subsequence $X_{n_j} \to X$ almost surely.

(4). This follows directly from Markov inequality. A counterexample is the escape to width infinity given in Example 1.5.3 of "M".

(5). See Exercise 1.5.8 of "M".

Prove by contradiction, assume that $X_n \not\to X$ in probability. Then we can find a subsequence X_{n_j} such that $\mathbf{P}(|X_{n_j} - X| > \varepsilon) \geq c$ for some $c > 0$ and every j. By assumption, that subsequence has a further subsequence $X_{n_{j_k}}$ that converges to X in probability. In particular, for sufficiently large k we must have $\mathbf{P}(|X_{n_{j_k}} - X| > \varepsilon) < c$, a contradiction.

(6). No, as a counterexample, let X_n be a sequence of random variables converging in probability but not almost surely (say the typewriter sequence), then any subsequence of it will also converges in probability, and thus yields a further subsequence converging almost surely by (3).

(7). It is easy to establish the claim by the fact that continuous functions of pointwise convergent sequences are pointwise convergent. Here we prove the claim by definitions given in these sets of note.

For convenience, we assume the X_n and X are real random variables. For any $\delta > 0$, define $B_\delta := \{x \in \mathbf{R} : \exists y \in \mathbf{R},\ |x - y| < \delta,\ |F(x) - F(y)| > \varepsilon\}$. If the event $|F(X_n) - F(X)| > \varepsilon$ holds, then either $|X_n - X| > \delta$ or $X \in B_\delta$. By subadditivity, this means

$\mathbf{P}(|F(X_n) - F(X)| > \varepsilon) \leq \mathbf{P}(|X_n - X| > \delta) + \mathbf{P}(X \in B_\delta)$.

By continuity of F (and continuity from above), sending $\delta \to 0$ makes the second term on the RHS vanish, by the hypothesis of $X_n \to X$ in probability, sending $n \to \infty$ makes the first term on the RHS vanish, so $F(X_n) \to F(X)$ in probability.

For the general case, we let $X_n := (X_n^{(1)}, \ldots, X_n^{(k)}),\ X := (X^{(1)}, \ldots, X^{(k)})$ and define B_δ similarly, then again

$\mathbf{P}(|F(X_n) - F(X)| > \varepsilon) \leq \mathbf{P}(|X_n - X| > \delta) + \mathbf{P}(X \in B_\delta)$.

The first term on the RHS equals $\mathbf{P}(\bigwedge_{i=1}^k (|X_n^{(i)} - X^{(i)}| > \frac{\delta}{\sqrt{k}}))$ vanishes as $n \to \infty$, and the second term on the RHS vanishes as $\delta \to 0$, so again we have $F(X_n) \to F(X)$ in probability.

(8). See Lemma 15 in note 1.

If $\mathbf{E}X_n$ diverges, then the conclusion is trivial, so we assume $\mathbf{E}X_n$ converges. By Proposition 6.6.6 of "Analysis", there is a subsequence X_{n_j} of X_n such that $\lim_j \mathbf{E}X_{n_j} = \liminf_n \mathbf{E}X_n$. By condition, there is a further subsequence $X_{n_{j_k}}$ of X_{n_j} such that $X_{n_{j_k}} \to X$ almost surely, and thus by Lemma 15 in note 1 $\mathbf{E}X \leq \liminf_k \mathbf{E}X_{n_{j_k}} = \lim_j \mathbf{E}X_{n_j} = \liminf_n \mathbf{E}X_n$.

(9). See Theorem 22 in note 1.

Solution 1: As $X_{n_j} \to X$ for some subsequence X_{n_j} and $|X_{n_j}| \leq Y$ almost surely for all j, $|X| \leq Y$ almost surely. Define $A_n := \{|X_n - X| > \varepsilon\}$, then $\mathbf{E}|X - X_n| \leq \int_{\Omega\setminus A_n} \varepsilon\ d\mathbf{P} + \int_{A_n} 2Y\ d\mathbf{P} \leq \varepsilon + 2\int_{A_n} Y\ d\mathbf{P}$.

To control $\int_{A_n} Y\ d\mathbf{P}$, we use monotone convergence on $\min(Y, n)$. Choose N large enough such that $\int_{A_n} Y - \min(Y, N)\ d\mathbf{P} < \varepsilon/2$, then $\int_{A_n} Y\ d\mathbf{P} < \varepsilon/2 + N\mathbf{P}(A_n) < \varepsilon$ for sufficiently large n with $\mathbf{P}(A_n) < \varepsilon/N$. Hence $\mathbf{E}|X - X_n| \leq 3\varepsilon$ and the claim follows from the triangle inequality.

Solution 2: By splitting into real and imaginary parts, we may assume without loss of generality that the X_n are real-valued. As Y is absolutely integrable, it is finite almost everywhere; after modification on a set of measure zero we may assume it is finite everywhere. From Fatou's lemma for convergence in probability applied to the unsigned functions $Y - X_n$ and $Y + X_n$, we have $\mathbf{E}(Y - X) \leq \liminf\limits_{n\to\infty} \mathbf{E}(Y - X_n)$ and $\mathbf{E}(Y + X) \leq \liminf_{n\to\infty} \mathbf{E}(Y + X_n)$.

Rearranging this (taking crucial advantage of the finite nature of the $\mathbf{E}Y$, and hence $\mathbf{E}X$ and $\mathbf{E}X_n$, we conclude that $\limsup\limits_{n\to\infty} \mathbf{E}X_n \leq \mathbf{E}X \leq \liminf\limits_{n\to\infty} \mathbf{E}X_n$, and the claim follows.

Solution 3: By condition, every subsequence X_{n_j} of X_n has a further subsequence $X_{n_{j_k}}$ converging to X almost surely. By Theorem 24 in note 1, $\lim_k \mathbf{E}X_{n_{j_k}} = \mathbf{E}X$. By the fact that a sequence a_n converges to L if and only if every subsequence has a further subsequence converging to L, the claim follows. □

Exercise 3

Proof. (1). Solution 1: It suffices to show that for any bounded continuous scalar functions F and G on the ranges of X and Y respectively, we have $\mathbf{E}F(X)G(Y) = (\mathbf{E}F(X))(\mathbf{E}G(Y))$ (as can be seen by specialising to those F, G that take values in $\{0, 1\}$).

By part (7) of Exercise 2, $F(X_n) \to F(X)$ in probability, and thus by dominated convergence in probability, $\mathbf{E}F(X)G(Y) = \lim_{n\to\infty} \mathbf{E}F(X_n)G(Y) = \lim_{n\to\infty}(\mathbf{E}F(X_n))(\mathbf{E}G(Y)) = \mathbf{E}F(X)\mathbf{E}G(Y)$, and we are done.

Solution 2: Without loss of generality, we may assume the X_n and X are real-valued. Since $X_n \xrightarrow{p} X$, there is a subsequence X_{n_j} such that $X_{n_j} \xrightarrow{a.s} X$, renumber the original sequence so that $X_n \xrightarrow{a.s} X$, we have:

$(X \leq t) = (\limsup_{n\to\infty} X_n \leq t) = \bigvee_{N\geq 1} \bigwedge_{n\geq N} (X_n \leq t)$.

In particular, we see that $(X \leq t) \in (\sigma(X_i))_{i\geq 1}$. As Y is independent of X_i for each i, $\sigma(Y)$ is independent of $\sigma(X_i)$ for each i, and since $(\sigma(X_i))_{i\geq 1}$ is generated by $\sigma(X_i)$ for $i = 1, 2, \ldots$, a simple application of the monotone class lemma shows that $\sigma(Y)$ is independent of $(\sigma(X_i))_{i\geq 1}$, and we are done.

Comment on solution 2: When we have a.s convergence of random variables, it could be easier sometimes to work with the σ-algebras generated by these variables, rather than directly with the variables themselves.

(2). As above, we assume that $X_n,\ X \in \mathbf{R}$. As $X_n \xrightarrow{d} X$, $X_{n_j} \xrightarrow{a.s} X$ for a subsequence X_{n_j} of X_n, so X is a real scalar tail random variable of X_{n_j}, the claim then follows from a Corollary of the Kolmogorov zero-one law. $\square$

Comment on the moment method:

1. Assuming a finite fourth moment $\mathbf{E}|X|^4 < \infty$, by the Hölder inequality with $p = q = 2$, we have $\mathbf{E}|X|^2 = \mathbf{E}|X|^2 \cdot 1 \leq (\mathbf{E}|X|^4)^{1/2} < \infty$.

Alternatively, let $Y_n := \min(|X|^2, n)$, since X is absolutely integrable, it is finite almost surely, so $|X|^2$ is finite almost surely and the non-decreasing sequence $Y_n \to |X|^2$ almost surely. By the monotone convergence theorem (or the horizontal truncation property), this implies that $\lim_n \mathbf{E}Y_n = \mathbf{E}|X|^2$. Apply the Jensen's inequality with $f(x) = x^2$ to Y_n (note that both Y_n and $f(Y_n)$ are now absolutely integrable), we see that $(\mathbf{E}Y_n)^2 \leq \mathbf{E}|Y_n|^2 \leq \mathbf{E}|X|^4 < \infty$, and in particular $\mathbf{E}Y_n < \infty$ for all n, giving $\mathbf{E}|X|^2 < \infty$.

2. Similarly, Hölder inequality gives

$|\mathbf{E}X_iX_jX_kX_l| \leq (\mathbf{E}|X_iX_j|^2)^{1/2}(\mathbf{E}|X_kX_l|^2)^{1/2}$. To bound the RHS, use Hölder inequality again:

$\mathbf{E}|X_iX_j|^2 = \mathbf{E}|X_i|^2|X_j|^2 \leq (\mathbf{E}|X_i|^4)^{1/2}(\mathbf{E}|X_j|^4)^{1/2} < \infty$.

Be cautious of the implication of integrability, as $\mathbf{E}|X| < \infty$, X is almost surely finite (almost surely bounded). However, if X is not integrable, it is not necessarily the case that $X = \infty$ on a positive measure set, non-integrability only implies non-negligible tails, i.e. $\mathbf{P}(|X| \geq M) > 0$ for any $M > 0$.

3. The number of quadruples (i, j, k, l) such that one is not distinct from the other three are $3n(n-1)$, with each contribute σ^4 to $\mathbf{E}|S_n|^4$.

4. The second display of proposition 6 follows form part (7) of Exercise 2: $\sqrt{\frac{X_1^2+\cdots+X_n^2}{n}}$ converges in probability to $\sqrt{1/3}$, so $\mathbf{P}(|\sqrt{\frac{X_1^2+\cdots+X_n^2}{n}} - \sqrt{1/3}| > \varepsilon\sqrt{1/3})$ goes to zero as $n \to \infty$ and the claim follows.

5. In the coupon collector problem, $X_i := T_{i,N} - T_{i-1,N}$ represents the time taken to collect the i^{th} distinct coupon after one has collected $i-1$ distinct coupons.

6. Part of the proof of proposition 8 should be changed to:

In order for the event $X_1 = j_1 \wedge \cdots \wedge X_N = j_N$ to hold, we must have $j_1 = 1$ and the first coupon Y_{j_1} can be arbitrary; then the coupons $Y_{j_1+1}, \ldots, Y_{j_1+j_2-1}$ have to be equal to Y_{j_1}, and $Y_{j_1+j_2}$ must be from one of the remaining $N-1$ elements of $\{1, \ldots, N\}$ not equal to Y_{j_1}; then $Y_{j_1+j_2+1}, \ldots, Y_{j_1+j_2+j_3-1}$ must be from the two-element set $\{Y_{j_1}, Y_{j_1+j_2}\}$, and $Y_{j_1+j_2+j_3}$ must be from the remaining $N-2$ elements of $\{1, \ldots, N\}$; and so on and so forth up to $Y_{j_1+\cdots+j_N}$.

7. Following proposition 8, Applying Chebyshev's inequality with the given bounds yields

$\mathbf{P}(|T_N - N\log N - O(N)| \geq \lambda N) = O(\lambda^{-2})$.

By triangle inequality, this implies $\mathbf{P}(|T_N - N\log N| - |O(N)| \geq \lambda N) = O(\lambda^{-2})$, so $\mathbf{P}(|T_N - N\log N| \geq (C+\lambda)N) = O(\lambda^{-2}) = O((\lambda+C)^{-2})$ for some constant C, set $\lambda := \lambda + C$.

Exercise 9

Proof. By definition, $\mathbf{E}X = \sum_{j=1}^{\infty} j(1-p)^{j-1}p = p\sum_{j=1}^{\infty} j(1-p)^{j-1} = p/p^2 = 1/p$. For simplicity, denote $1-p$ by q, and note that:

$$\begin{aligned}
\mathbf{E}X^2 &= \sum_{j=1}^{\infty} j^2 q^{j-1} p = \sum_{j=1}^{\infty} (j-1+1)^2 q^{j-1} p \\
&= \sum_{j=1}^{\infty} (j-1)^2 q^{j-1} p + 2\sum_{j=1}^{\infty} (j-1) q^{j-1} p + \sum_{j=1}^{\infty} q^{j-1} p \\
&= \sum_{i=0}^{\infty} i^2 q^i p + 2\sum_{i=0}^{\infty} i q^i p + 1 \\
&= q\mathbf{E}X^2 + 2q\mathbf{E}X + 1
\end{aligned}$$

Using the fact that $\mathbf{E}X = 1/p$, we solve for $\mathbf{E}X^2$ to get $\mathbf{E}X^2 = \frac{2-p}{p^2}$, so $\mathbf{Var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}$. $\square$

The proof of the variance of the geometric distribution presented here is an example of the problem-solving principle of *introducing something new*.

Exercise 10

Proof. Solution 1: Let $S_n := \sum_{i=1}^n 1_{E_i}$, we use the Paley-Zygmund inequality (the second moment method).

As the sequence S_n is non-decreasing, $\lim_n S_n = \limsup_{n\to\infty} S_n$, and it suffices to show that $\mathbf{P}(\limsup_{n\to\infty} S_n = \infty) = 1$.

Fix some large $M > 0$. For any $0 \le \theta \le 1$ small, let $n' = n(\theta)$ be sufficiently large with $\theta \mathbf{E} S_{n'} > M$. For all $n \ge n'$, the Paley-Zygmund inequality gives $\mathbf{P}(S_n > M) \ge \mathbf{P}(S_n > \theta \mathbf{E} S_n) \ge (1-\theta)^2 \frac{(\mathbf{E}S_n)^2}{(\mathbf{E}|S_n|^2)} = (1-\theta)^2$ (by the independence hypothesis, $(\mathbf{E}S_n)^2 = \mathbf{E}|S_n|^2$). In particular, $\mathbf{P}(S_n > M) \to 1$.

We thus obtain $\mathbf{P}(\limsup_{n\to\infty} S_n > M) = \lim_{N\to\infty} \mathbf{P}(\bigvee_{n\ge N}(S_n > M)) \ge \lim_{N\to\infty} \mathbf{P}(S_N > M) = 1$, since M is arbitrary, the claim follows from continuity from above.

Solution 2: We calculate the mean $\mathbf{E}S_n = \sum_{i=1}^n \mathbf{P}(E_i)$ and the variance $\mathbf{Var}(S_n) = \sum_{i=1}^n \mathbf{P}(E_i)(1-\mathbf{P}(E_i))$ (where we use the independence hypothesis). By the Chebyshev's inequality:

$$\mathbf{P}(|S_n - \mathbf{E}S_n| \ge \frac{1}{2}\mathbf{E}S_n) \le \frac{4\mathbf{Var(S_n)}}{(\mathbf{E}S_n)^2} \le \frac{4\sum_{i=1}^n \mathbf{P}(E_i)(1-\mathbf{P}(E_i))}{(\sum_{i=1}^n \mathbf{P}(E_i))^2} \quad \le \frac{4}{\sum_{i=1}^n \mathbf{P}(E_i)} \to 0$$

So in particular $\mathbf{P}(|S_n - \mathbf{E}S_n| \ge \frac{1}{2}\mathbf{E}S_n) \to 0$, or equivalently

$\mathbf{P}(|S_n - \mathbf{E}S_n| < \dfrac{1}{2}\mathbf{E}S_n) \to 1$, implying that $\mathbf{P}(S_n > \dfrac{\mathbf{E}S_n}{2}) \to 1$.

For any large $M > 0$, we can pick $n' > 0$ sufficiently large such that $\frac{\mathbf{E}S_n}{2} > M$, the rest is the same as above. □

Exercise 11

Proof. Define the events $E_n := \bigwedge_{i=(n-1)k+1}^{nk}(X_i = a_{i \bmod k})$, so $E_1 = \bigwedge_{i=1}^{k}(X_i = a_i)$, $E_2 = \bigwedge_{i=k+1}^{2k}(X_i = a_{i \bmod k}), \ldots$ etc. These events are all independent, and $\sum_{n=1}^\infty \mathbf{P}(E_n) = \sum_{n=1}^\infty 1/N^k = \infty$ (assuming $|A| = N$), so by the second Borel-Cantelli lemma, almost surely an infinite number of the E_n hold simultaneously. That is, the finite word $a_1 \ldots a_k$ appears infinitely often in the string $X_1X_2X_3\ldots$ almost surely. □

Exercise 12

Proof. As in the case of a single sequence, we normalise the mean μ to equal zero, by replacing each $X_{i,n}$ with $X_{i,n} - \mu$, so that S_n gets replaced by $S_n - \mu n$ (and S_n/n by $S_n/n - \mu$). To prove the strong or weak law of large numbers, it then suffices to do so in the mean zero case $\mu = 0$.

(1). Let $\alpha := \sup_{i,n} \mathbf{E}|X_{i,n}|^2$. The first moment calculation then shows that S_n has mean zero. Now we compute the variance of S_n, which in the mean zero case is simply $\mathbf{E}|S_n|^2$; note from the triangle inequality that this quantity is finite. By linearity of expectation, we have

$$\mathbf{Var}(S_n) = \mathbf{E}|X_{i,1} + \cdots + X_{i,n}|^2 = \sum_{1 \le i,j \le n} \mathbf{E} X_{i,n} X_{j,n}.$$

(All expressions here are absolutely integrable thanks to the Cauchy-Schwarz inequality.) If $i = j$, then the term $\mathbf{E} X_{i,n} X_{j,n}$ is equal to $\mathbf{E}|X_{i,n}|^2$. If $i \neq j$, then by hypothesis $X_{i,n}$ and $X_{j,n}$ are independent and mean zero, and thus $\mathbf{E} X_{i,n} X_{j,n} = (\mathbf{E} X_{i,n})(\mathbf{E} X_{j,n}) = 0$.

Putting all this together, we obtain

$$\mathbf{Var}(S_n) = \sum_{i=1}^{n} \mathbf{E}|X_{i,n}|^2 \le n\alpha, \text{ or equivalently } \mathbf{Var}(S_n/n) \le \frac{\alpha}{n}.$$

This bound was established in the mean zero case, but it is clear that it also holds in general, since subtracting a constant from a random variable does not affect its variance.

Insert this variance bound into Chebyshev's inequality, we obtain the bound

$$\mathbf{P}(|\frac{S_n}{n} - \mu| \ge \varepsilon) \le \frac{1}{n}\frac{\alpha}{\varepsilon^2}$$

for any natural number n and $\varepsilon > 0$, whenever the $X_{i,n}$ all have mean μ and bounded variance. The right-hand side goes to zero as $n \to \infty$ for fixed ε, so we have in fact established the weak law of large numbers for triangular arrays in the case that the $X_{i,n}$ have bounded variance.

(2). We can expand

$$\mathbf{E}|S_n|^4 = \mathbf{E}|X_{1,n} + \cdots + X_{n,n}|^4 = \sum_{1 \le i,j,k,l \le n} \mathbf{E} X_{i,n} X_{j,n} X_{k,n} X_{l,n}.$$

Note that all expectations here are absolutely integrable by Hölder's inequality and the hypothesis of bounded fourth moment. Suppose for instance that i is distinct from j, k, l, then $X_{i,n}$ is independent of $(X_{j,n}, X_{k,n}, X_{l,n})$ (even if some of the j, k, l are equal to each other) and so

$\mathbf{E} X_{i,n} X_{j,n} X_{k,n} X_{l,n} = (\mathbf{E} X_{i,n})(\mathbf{E} X_{j,n} X_{k,n} X_{l,n}) = 0$ since $\mathbf{E} X_{i,n} = \mu = 0$. Similarly for permutations. This leaves only a few quadruples (i, j, k, l) for which $\mathbf{E} X_{i,n} X_{j,n} X_{k,n} X_{,nl}$ could be non-zero: the three cases $i = j \neq k = l$, $i = k \neq j = l$, $i = l \neq j = k$ where each of the indices i, j, k, l is paired up with exactly one other index; and the diagonal case $i = j = k = l$. If for instance $i = j \neq k = l$, then

$$\mathbf{E} X_{i,n} X_{j,n} X_{k,n} X_{l,n} = \mathbf{E} X_{i,n}^2 X_{k,n}^2 = (\mathbf{E} X_{i,n}^2)(\mathbf{E} X_{k,n}^2) \le \alpha^2.$$

Similarly for the cases $i = k \neq j = l$ and $i = l \neq j = k$, which gives a total contribution of at most $3n(n-1)\alpha^2$ to $\mathbf{E}|S_n|^4$. Finally, when $i = j = k = l$, then $\mathbf{E}X_{i,n}X_{j,n}X_{k,n}X_{l,n} = \mathbf{E}X_{i,n}^4$, and there are n contributions of this form to $\mathbf{E}|S_n|^4$. We conclude that

$\mathbf{E}|S_n|^4 \leq 3n(n-1)\alpha^2 + n\mathbf{E}|X_{i,n}|^4$

and hence by Markov's inequality $\mathbf{P}(|\frac{S_n}{n}| > \varepsilon) \leq \frac{3n(n-1)\alpha^2 + n\mathbf{E}|X_{i,n}|^4}{\varepsilon^4 n^4}$ for any $\varepsilon > 0$. If we remove the normalisation $\mu = 0$, we conclude that

$$\mathbf{P}(|\frac{S_n}{n} - \mu| > \varepsilon) \leq \frac{3n(n-1)\alpha^2 + n\mathbf{E}|X_{i,n} - \mu|^4}{\varepsilon^4 n^4}$$

The right-hand side decays like $O(1/n^2)$, which is now summable in n. Thus we may now apply the Borel-Cantelli lemma and conclude the strong law of large numbers in the case when one has bounded fourth moment $\sup_{i,n} \mathbf{E}|X_{i,n}|^4 < \infty$. $\square$

Exercise 13

Proof. (1). For any unordered pairs $\{i, j\} \in V_n$, we define the random variables $X_{ij,\binom{n}{2}} := 1_{\{i,j\} \in E_n}$, it can be seen from condition that $\mathbf{E}X_{ij,\binom{n}{2}} = 1/2$ for all i, j and the $X_{ij,\binom{n}{2}}$ are independent. Clearly, we have $\sup_{i,j,n} \mathbf{E}|X_{ij,\binom{n}{2}}|^4$, by the strong law for triangular arrays, $|E_n|/\binom{n}{2} \to 1/2$ almost surely.

(2). Likewise, we define the random variables $X_{ijk,\binom{n}{3}} := 1_{\{i,j\},\{j,k\},\{i,k\} \in E_n}$ for all unordered triples $\{i, j, k\}$ in V_n, and normalise the mean $\mathbf{E}X_{ijk,\binom{n}{3}} = 0$ for all such triples, by replacing each $X_{ijk,\binom{n}{3}}$ with $X_{ijk,\binom{n}{3}} - 1/8$, so that $|T_n|$ gets replaced by $|T_n| - \frac{1}{8}\binom{n}{3}$ (and $|T_n|/\binom{n}{3}$ by $|T_n|/\binom{n}{3} - 1/8$). To prove the given claim, it then suffices to do so in this mean-zero setting.

The first moment calculation shows that $|T_n|$ has mean zero. Hence

$$\mathbf{Var}(|T_n|) = \mathbf{E}|T_n|^2 = \mathbf{E}|\sum_{\substack{\{i,j,k\} \in V_n \\ \text{unordered}}} X_{ijk,\binom{n}{3}}|^2 = \sum_{1 \leq s,t \leq \binom{n}{3}} \mathbf{E}X_s X_t,$$

where we relabel the $X_{ijk,\binom{n}{3}} = X_s$ to range over the set $\{1, \ldots, \binom{n}{3}\}$. Note that for X_s and X_t to be correlated, their corresponding triples need to share at least one edge, so the RHS of the above identity equals

$$\binom{n}{2} \sum_{1 \leq k,l \leq n-2} \mathbf{E}X_{ijk,\binom{n}{3}} X_{ijl,\binom{n}{3}} = \binom{n}{2}[\binom{n-2}{2}/2^5 + (n-2)/2^3]$$

This bound was established in the mean zero case, but it is clearly that it holds in general, since subtracting a constant from a random variable does not affect its variance.

Insert this bound into Chebyshev's inequality, we obtain the bound $\mathbf{P}(||T_n|/\binom{n}{3} - 1/8| \geq \varepsilon) \leq O(\frac{1}{n^2})/\varepsilon^2$

for any natural number n and any $\varepsilon > 0$, whenever the $X_{ijk,\binom{n}{3}}$ all have mean 1/8 and bounded variance. The right-hand side goes to zero as $n \to \infty$ for fixed ε, so $|T_n|/\binom{n}{3} \to 1/8$ in probability.

(3). The bound in the last part $\mathbf{P}(||T_n|/\binom{n}{3} - 1/8| \geq \varepsilon) \leq O(\frac{1}{n^2})/\varepsilon^2$ is summable in n, so we may apply the Borel-Cantelli lemma to get the claim. □

Exercise 14

Proof. (1). By definition, $(A_nA_n^*)_{ii} = n$ for all $1 \leq i \leq n$, so $\mathrm{tr}A_nA_n^*/n^2 = n^2/n^2 = 1$ deterministically.

(2). When $k = 1$, $\mathrm{tr}(A_nA_n^*)^k = n^2$ deterministically and $\mathbf{E}\mathrm{tr}(A_nA_n^*)^k/n^{k+1} = 1$ for all n. Let $r_i(A)$ denote the ith row of the matrix A, then $r_i(A_nA_n^*) = (\sum_{1\leq t\leq n} a_{it}a_{1t}, \ldots, \sum_{1\leq t\leq n} a_{it}a_{nt})$, and we have:

$$\begin{aligned}(A_nA_n^*)^2_{ii} &= r_i(A_nA_n^*) \cdot r_i(A_nA_n^*) \\ &= \sum_{1\leq t_1,t_2\leq n} a_{it_1}a_{1t_1}a_{it_2}a_{1t_2} + \cdots + \sum_{1\leq t_1,t_2\leq n} a_{it_1}a_{nt_1}a_{it_2}a_{nt_2}.\end{aligned}$$

Taking expectation, each sum contributes $O(n)$ (the indices need to be paired up in order to have a non-zero contribution), and there are n sums. This implies that $\mathbf{E}(A_nA_n^*)^2_{ii} = O(n^2)$, and thus $\mathbf{E}\mathrm{tr}(A_nA_n^*)^2 = O(n^3)$. By a similar computation,

$$\begin{aligned}(A_nA_n^*)^3_{ii} &= r_i(A_nA_n^*)^2 \cdot r_i(A_nA_n^*) \\ &= \sum_{t_1,t_2,t_3,t_4} a_{it_1}a_{t_3t_1}a_{1t_2}a_{t_3t_2}a_{it_4}a_{1t_4} + \cdots + \sum_{t_1,t_2,t_3,t_4} a_{it_1}a_{t_3t_1}a_{nt_2}a_{t_3t_2}a_{it_4}a_{nt_4}.\end{aligned}$$

From which we get $\mathbf{E}(A_nA_n^*)^3_{ii} = O(n^3)$, and thus $\mathbf{E}\mathrm{tr}(A_nA_n^*)^3 = O(n^4)$.

In general, by the nature of dot product, each time one raises the power k by 1, the sums in $r_i(A_nA_n^*)^k$ obtain 2 additional indices. i.e., $r_i(A_nA_n^*)^k = (\sum_{t_1,\ldots,t_{2k-1}}, \ldots, \sum_{t_1,\ldots,t_{2k-1}})$, it follows that $(A_nA_n^*)^{k+1}_{ii} = r_i(A_nA_n^*)^k \cdot r_i(A_nA_n^*) = \sum_{t_1,\ldots,t_{2k}} + \cdots + \sum_{t_1,\ldots,t_{2k}}$. Taking expectation, it becomes $O(n^{k+1})$.

Consequently, we have $\mathbf{E}\mathrm{tr}(A_n A_n^*)^k/n^{k+1} = O(n^{k+1})/n^{k+1} = O_k(1)$, as desired.

(3). Fix some $\varepsilon, \delta > 0$. As $A_n A_n^*$ is symmetric, the operator norm $\|A_n A_n^*\|_{op}$ is equal to the ℓ^∞ norm $\|A_n A_n^*\|_{op} = \max\limits_{1 \leq i \leq n} \lambda_i$ of the eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$ of $A_n A_n^*$ (by Corollary 1 of Theorem 6.43 of "L", these values are non-negative). Also, we have the standard linear algebra identity $\mathrm{tr}(A_n A_n^*)^k = \sum\limits_{i=1}^{n} \lambda_i^k$. Consequently, for any natural number k, we have the inequalities

$$\|A_n\|_{op}^{2k} = \|A_n A_n^*\|_{op}^k \leq \mathrm{tr}(A_n A_n^*)^k \leq n\|A_n A_n^*\|_{op}^k = n\|A_n\|_{op}^{2k}$$

It follows that for any natural number k,
$\|A_n\|_{op}/n^{1/2+\varepsilon} \leq [\mathrm{tr}(A_n A_n^*)^k]^{\frac{1}{2k}}/n^{1/2+\varepsilon}$
and
$\|A_n\|_{op}/n^{1/2-\varepsilon} \geq [\mathrm{tr}(A_n A_n^*)^k]^{\frac{1}{2k}}/n^{\frac{1}{2k}+\frac{1}{2}-\varepsilon}$.

Let E_n be the event $[\mathrm{tr}(A_n A_n^*)^k]^{\frac{1}{2k}}/n^{1/2+\varepsilon} > \delta$, which is equivalent to the event $\mathrm{tr}(A_n A_n^*)^k/n^{k(1+2\varepsilon)} > \delta$, where we re-scale $\delta := \delta^{2k}$. By part (2) and the Markov inequality, we obtain the bound

$$\mathbf{P}(E_n) \leq \frac{C_k}{\delta n^{2\varepsilon k-1}}.$$

for all k. Choose sufficiently large k (greater than 2) depending on ε such that the denominator is $O(n^2)$ (say), then $\mathbf{P}(E_n)$ and hence $\mathbf{P}(\|A_n\|_{op}/n^{1/2+\varepsilon} > \delta)$ are summable in n. By the Borel-Cantelli lemma, this implies that $\|A_n\|_{op}/n^{1/2+\varepsilon}$ converges almost surely to zero.

Finally, setting $k = 1$ in the second inequality gives the bound

$$\|A_n\|_{op}/n^{1/2-\varepsilon} \geq n^{\varepsilon},$$

which implies that $\|A_n\|_{op}/n^{1/2-\varepsilon}$ diverges almost surely to infinity. □

Exercise 15

Proof. (1). For all $n > 2$, let $X_n = 1_{\mathcal{P}}(n)$ be independent Bernoulli random variables with parameters $1/\log n$. Take x to be a natural number, and let $(Y_{n,x})_{n,x>2:n\leq x}$ be the triangular array given by $Y_{n,x} := \frac{1}{x/\log x} X_n$.

Consider the partial sums $S_x = \frac{\sum_{2<n\leq x} X_n}{x/\log x} = Y_{3,x} + \cdots + Y_{x,x}$. For convenience, we normalise each X_n to have zero mean, by replacing X_n by $X_n - 1/\log n$, so that S_x also gets replaced by

$$S_x - \frac{\sum_{2<n\le x} 1/\log n}{x/\log x} = S_x - \frac{\int_3^x dt/\log t + O(1)}{x/\log x} = S_x - (1 + O(1/\log x)),$$

where we use integration by parts for the last inequality. We can then expand

$$\mathbf{E}|S_x|^4 = \mathbf{E}|Y_{3,x} + \cdots + Y_{x,x}|^4 = \sum_{2<i,j,k,l\le x} \mathbf{E}|Y_{i,x}Y_{j,x}Y_{k,l}Y_{l,x}|^4.$$

The independence hypothesis leaves only a few quadruples (i,j,k,l) for which $\mathbf{E}Y_{i,x}Y_{j,x}Y_{k,x}Y_{l,x}$ could be non-zero: the three cases $i=j\neq k=l, i=k\neq j=l$, $i=l\neq j=k$ where each of the indices i,j,k,l is paired up with exactly one other index; and the diagonal case $i=j=k=l$. If for instance $i=j\neq k=l$, then
$$\mathbf{E}Y_{i,x}Y_{j,x}Y_{k,x}Y_{l,x} = \mathbf{E}Y_{i,x}^2Y_{k,x}^2 = (\mathbf{E}Y_{i,x}^2)(\mathbf{E}Y_{k,x}^2) \le (\frac{\log x}{x})^2 \times (\frac{\log x}{x})^2.$$

Similarly for the cases $i=k\neq j=l$ and $i=l\neq j=k$, which gives a total contribution of at most $3x(x-1)(\frac{\log x}{x})^4$ to $\mathbf{E}|S_x|^4$. Finally, when $i=j=k=l$, then $\mathbf{E}Y_{i,x}Y_{j,x}Y_{k,x}Y_{l,x} \le 2(\frac{\log x}{x})^4$, and there are at most x contributions of this form to $\mathbf{E}|S_x|^4$. We conclude that
$$\mathbf{E}|S_x|^4 \le 3x(x-1)(\frac{\log x}{x})^4 + 2x(\frac{\log x}{x})^4 = \frac{(3x-1)\log^4 x}{x^3}$$

and hence by Markov's inequality,
$$\mathbf{P}(|S_x| > \varepsilon) \le \frac{(3x-1)\log^4 x}{\varepsilon^4 x^3} \text{ for any } \varepsilon > 0.$$

Remove the normalisation, we conclude that
$$\mathbf{P}(|S_x - (1 + O(1/\log x))| > \varepsilon) \le \frac{(3x-1)\log^4 x}{\varepsilon^4 x^3} \text{ for any } \varepsilon > 0.$$

By the comparison test, the right-hand side is summable in x. The Borel-Cantelli lemma thus implies that S_x converges almost surely to one as $x\to\infty$, as desired.

(2). Let X_n be as in part (1), and $(Y_{n,x})_{n,x>2:n\le x}$ be the triangular array given by $Y_{n,x} := \frac{X_n - 1/\log n}{x^{1/2+\varepsilon}}$, the $Y_{n,x}$ have mean zero and are independent row-wise, thus
$$\mathbf{E}(\sum_{2<n\le x} Y_{n,x})^k = \frac{1}{x^{k/2+k\varepsilon}}\mathbf{E}(\sum_{2<n\le x} X_n - 1/\log n)^k = O(x^{-k\varepsilon})$$
for any fixed even number k (the only terms in $(\sum_{2<n\le x} X_n - 1/\log n)^k$ that give a non-zero contribution are those in which each $Z_n = X_n - 1/\log n$ appears at least twice, so there are at most $k/2$ distinct indices of n that arise, and so there are only $O_k(x^{k/2})$ such terms, each contributing $O(1)$). By Markov's inequality, $\mathbf{P}(|\sum_{2<n\le x} Y_{n,x}| > \delta) \le O(1/\delta^k x^{k\varepsilon})$ for any $\delta > 0$ and k even. Setting

k sufficiently large depending on ε, we see the LHS is summable in x, and the claim follows from the fact that $\sum_{2\le n\le x} 1/\log n = \mathrm{Li}(x) + O(1)$.

Or, one can also utilize a more powerful concentration inequality, such as Hoeffding's inequality.

(3). See Prediction 8 in note 254A, Supplement 4.

We show that $|\{n \le x : n, n+2 \in \mathcal{P}\}| = (1+o(1))\dfrac{x}{\log^2 x}$ almost surely as $x \to \infty$, by which the claim follows. By the Borel-Cantelli lemma again, it suffices to show that for each $x > 2$, we have the above estimate with probability $1 - O(x^{-2+o(1)})$ (say) as $x \to \infty$.

Fix x. We write the left-hand side as

$$O(1) + \sum_{2\le n\le x} \frac{1}{(\log n)\log(n+2)} + X_n,$$

where $X_n := 1_{\mathcal{P}'}(n)1_{\mathcal{P}'}(n+2) - \dfrac{1}{(\log n)\log(n+2)}$.

To estimate the sum $\sum_{2\le n\le x} \frac{1}{(\log n)\log(n+2)}$, we can easily check that the contribution with $n \le x/\log^3 x$ is $o(x/\log^2 x)$, and for the remaining portion we have $\frac{1}{(\log n)\log(n+2)} = \frac{1+o(1)}{\log^2 x}$ (let $f = \frac{1}{(\log n)\log(n+2)}$ where $n = x/\log^3 x$, then $f/(1/\log^2 x) \to 1$ as $x \to \infty$), and thus

$$\sum_{2\le n\le x} \frac{1}{(\log n)\log(n+2)} = (1+o(1))\frac{x}{\log^2 x}.$$

It thus suffices to show that

$\displaystyle\sum_{2\le n\le x} X_n = o(\frac{x}{\log^2 x})$ with probability $1 - O(x^{-2+o(1)})$. But from construction, each X_n has mean zero, is bounded by $O(1)$, and X_n is independent of X_m unless $|n-m| \le 2$ (distance 1 apart incurs a dependence as there are no consecutive primes except for 2 and 3), from which we easily calculate that $\mathbf{E}(\displaystyle\sum_{2\le n\le x} X_n)^4 = O(x^2)$. A direct application of Markov's inequality then gives the required bound with the desired failure probability.

(4). Take the r.v $\sum_{3\le n\le x} 1_{\mathcal{P}}(n)1_{\mathcal{P}}(x-n)$, which we interpret as the number of ways x (taken to be a natural number) is expressible as the sum of two Cramér random model primes, scaled by a factor of two. Consider only integers greater than 2 in $\mathcal{P}$, by independence:

$$\mathbf{E}S_x = \sum_{3\le n\le x-3} \frac{1}{\log n \log(x-n)} \sim 2\int_3^{x/2} \frac{dt}{\log t \log(x-t)} := 2I_x.$$

Moreover, since $\displaystyle\frac{2}{\log(x-3)}\int_3^{x/2}\frac{dt}{\log t} \le 2I_x \le \frac{2}{\log x - \log 2}\int_3^{x/2}\frac{dt}{\log t}$,

one may deduce from $\displaystyle\int_3^x \frac{dt}{\log t} \sim x/\log x$ that $\mathbf{E}S_x \sim x/\log^2 x$.

We normalise each $X_n = 1_{\mathcal{P}}(n)1_{\mathcal{P}}(x-n)$ to have mean zero by replacing it with $Y_n := 1_{\mathcal{P}}(n)1_{\mathcal{P}}(x-n) - \dfrac{1}{\log n \log(x-n)}$, so that S_x also gets replaced by $S'_x = \displaystyle\sum_{3\leq n\leq x-3} Y_n$. As the Y_n are independent, have mean zero and of size $O(1)$, one can calculate the fourth moment $\mathbf{E}(\sum_{3\leq n\leq x-3} Y_n)^4 = O(x^2)$. By Markov's inequality, this implies $\forall\varepsilon > 0$,

$$\mathbf{P}(|\frac{S'_x}{x/\log^2 x}| > \varepsilon) \leq \frac{\log^8 x O(x^2)}{\varepsilon^4 x^4} = O(\frac{\log^8 x}{x^2}),$$

which is summable in x by the comparison test. By the Borel-Cantelli lemma, $\dfrac{S'_x}{x/\log^2 x}$ thus converges almost surely to zero. Equivalently, we have

$\mathbf{P}(\displaystyle\lim_{x\to\infty} \frac{S_x}{x/\log^2 x} = 1) = 1$. i.e, $S_x \overset{\text{a.s}}{\sim} x/\log^2 x$, and one obtains the claim.

Alternatively, one relatively easy way to proceed here is to locate a number of independent events that would produce twins, and apply the converse of the Borel-Cantelli lemma for independent events. $\square$

Exercise 16

Proof. (1). As n is chosen uniformly at random between 1 and x, the probability of a given prime p with $p \leq x^{1/10}$ being a factor of n is $\dfrac{\lfloor x/p \rfloor}{x} = \dfrac{1}{p} + O(1/x)$, hence $\mathbf{E}1_{p|n} = \dfrac{1}{p} + O(1/x)$. By linearity of expectation and Mertens' theorem,

$$\mathbf{E}(\sum_{p\leq x^{1/10}} 1_{p|n}) = \sum_{p\leq x^{1/10}} 1/p + O(1/x) = \log\log x + O(1).$$

Similarly, $\mathrm{Var}(1_{p|n}) = \dfrac{1}{p}(1-\dfrac{1}{p}) + O(1/x)$, and if $p \neq q$ are primes, $p|n$ and $q|n$ if and only if $pq|n$, so

$$\begin{aligned}\mathrm{Cov}(1_{p|n}, 1_{q|n}) &= \mathbf{E}(1_{p|n}1_{q|n}) - \mathbf{E}(1_{p|n})\mathbf{E}(1_{q|n}) = \mathbf{E}(1_{pq|n}) - \mathbf{E}(1_{p|n})\mathbf{E}(1_{q|n}) \\ &= \frac{\lfloor x/pq \rfloor}{x} - \frac{\lfloor x/p \rfloor}{x}\frac{\lfloor x/q \rfloor}{x} = \frac{1}{pq} + O(1/x) - [\frac{1}{p} + O(1/x)][\frac{1}{q} + O(1/x)] \\ &\leq \frac{1}{pq} + O(1/x) - \frac{1}{pq} = O(1/x)\end{aligned}$$

Expanding the formula $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$, we see

$$\begin{aligned}\mathrm{Var}(\sum_{p\le x^{1/10}} 1_{p|n}) &= \sum_{p\le x^{1/10}} \mathrm{Var}(1_{p|n}) + \sum_{p\neq q} \mathrm{Cov}(1_{p|n}, 1_{q|n}) \\ &= \log\log x + O(1) + 2x^{1/10} O(1/x) \\ &= O(\log\log x)\end{aligned}$$

(2). As no $n \le x$ can have more than ten prime factors larger than $x^{1/10}$, one has $\omega(n) - 10 \le \sum_{p\le x^{1/10}} 1_{p|n} \le \omega(n)$, so that large deviation bounds for $X = \sum_{p\le x^{1/10}} 1_{p|n}$ will translate into asymptotically similar bounds for ω. (Here 10 could be any large constant.)

Insert the bounds in part (1) into Chebyshev's inequality, one gets

$$\mathbf{P}(|\frac{X - \log\log x}{g(x)\sqrt{\log\log x}}| > \varepsilon) = \mathbf{P}(|X - \log\log x| > \varepsilon|g(x)|\sqrt{\log\log x}) \le \frac{O(1)}{\varepsilon^2 g^2(x)}.$$

As $|X - \omega| \le 10$ the same holds for ω and we are done. $\square$

Exercise 17

Proof. (1). Clearly, $\mathbf{H}(X) \ge 0$ from definition. Since the log function is concave, by Jensen's inequality (flipped for concave f) one has $\mathbf{H}(X) = \mathbf{E}\log\frac{1}{\mathbf{P}(X)} \le \log\mathbf{E}\frac{1}{\mathbf{P}(X)} = \log|A|$. $\mathbf{H}(X) = 0$ holds when the distribution of X is concentrated in a single point of A, i.e, $\exists x_0 \in A$ such that $\mathbf{P}(X = x_0) = 1$, while $\mathbf{H}(X) = \log|A|$ holds when X is distributed uniformly on A.

(2). Applying the weak law of large numbers on the sequence of random variables $\log\frac{1}{\mathbf{P}(X_i)}$, one obtains $\mathbf{P}(|\log(\prod_{i=1}^{n}\frac{1}{\mathbf{P}(X_i)}) - n\mathbf{H}(X)| > n\varepsilon) \xrightarrow{n} 0$, and in particular $\mathbf{P}(|\log(\prod_{i=1}^{n}\frac{1}{\mathbf{P}(X_i)}) - n\mathbf{H}(X)| > n\mathbf{H}(X)\varepsilon) \xrightarrow{n} 0$, but this quantity is precisely the proportion of A^n that lies outside of Ω, giving the first claim.

By the first claim, for sufficiently large n, one has

$1-\varepsilon \leq \sum_{\vec{x}\in\Omega} \mu_{\vec{X}}(\{\vec{x}\}) = \mathbf{P}(\vec{X}\in\Omega) \leq 1.$

From the definition of Ω, it then follows that $(1-\varepsilon)\exp((1-\varepsilon)n\mathbf{H}(X)) \leq |\Omega| \leq \exp((1+\varepsilon)n\mathbf{H}(X))$. As $1-\varepsilon \geq \exp(-\varepsilon n\mathbf{H}(X))$ for large n, we get the desired bound on the size of Ω. □

Exercise 18

Proof. Fix some $M > 0$. For any natural number m, we truncate X by $X^m := \min(X, m)$, and define $S_n^m = X_1^m + \cdots + X_n^m$ similarly. By the monotone convergence theorem and the non-integrability of X, we can choose sufficiently large m such that $\mathbf{E}X^m > M$. Then we have
$\mathbf{P}(S_n/n \geq M) \geq \mathbf{P}(S_n^m/n \geq M) \geq 1 - \mathbf{P}(|S_n^m/n - \mathbf{E}X^m| > \varepsilon)$
for any $\varepsilon > 0$ sufficiently small (while satisfying $\mathbf{E}X^m > M + \varepsilon$). By the weak law of large numbers, the RHS tends to 1 as $n \to \infty$, giving the claim. □

Exercise 19

Proof. As the hint suggests, it suffices to show that $X_n/n\log_2 n$ is almost surely unbounded. For the Petersburg random variable X with $b_n = n\log_2 n$, note that $\mathbf{P}(X > a) \geq \dfrac{1}{a}$ for any $a \geq 1$. (If k is the integer such that $2^k \leq a < 2^{k+1}$, then the given probability equals $\dfrac{1}{2^k} \geq \dfrac{1}{a}$) Therefore for any constant $M > 1$ and $n \geq 2$, $\mathbf{P}(\dfrac{X_n}{b_n} > M) = \mathbf{P}(X_n > Mb_n) \geq \dfrac{1}{Mb_n}$. In particular, $\sum_n \mathbf{P}(\dfrac{X_n}{b_n} > M) = \infty$, and the second Borel-Cantelli lemma implies that $\mathbf{P}(\limsup_{n\to\infty} \dfrac{X_n}{b_n} = \infty) = 1$. By definition, it also suggests that $\mathbf{P}(\lim_{n\to\infty} \dfrac{S_n}{b_n} = \infty) = 1$, as desired. □

Exercise 20

Proof. (1). We have $\int_{\mathbf{R}} \dfrac{1}{\pi}\dfrac{1}{1+x^2}\ dx = 1/\pi \tan^{-1} x\Big|_{-\infty}^{\infty} = 1$. i.e. the Cauchy distribution exists.

(2). Let X be a Cauchy random variable, by Exercise 36 of note 1 and integration by parts, $\mathbf{E}|X| = 2\int_0^{\infty} \dfrac{1}{\pi}\dfrac{x}{1+x^2}\ dx = \infty$, so X is not absolutely integrable.

(3). Let $S_n := |X_1| + \cdots + |X_n|$. We truncate each $|X_i|$ for $1 \leq i \leq n$ at Cn for some C to be chosen later, by writing $|X_i| = |X_i|_{\leq Cn} + |X_i|_{>Cn}$, where $|X_i|_{\leq Cn} := |X_i| 1_{|X_i| \leq Cn}$ and $|X_i|_{>Cn} := |X_i| 1_{|X_i|>Cn}$, similarly decompose $S_n = S_{n,\leq} + S_{n,>}$, where $S_{n,\leq} := |X_1|_{\leq Cn} + \cdots + |X_n|_{\leq Cn}$ and $S_{n,>} := |X_1|_{>Cn} + \cdots + |X_n|_{>Cn}$.

The random variable $|X| 1_{|X| \leq Cn}$ can be computed to have mean

$$\mathbf{E}(|X| 1_{|X| \leq Cn}) = \frac{1}{\pi} \int_{\mathbf{R}} \frac{|x| 1_{|x| \leq Cn}}{1+x^2}\, dx = \frac{1}{\pi} \int_{-Cn}^{Cn} \frac{|x|}{1+x^2}\, dx = \log(C^2 n^2 + 1)/\pi$$

and we can upper bound the variance by

$$\mathbf{Var}(|X| 1_{|X| \leq Cn}) \leq \mathbf{E}(|X| 1_{|X| \leq Cn})^2 = \frac{1}{\pi} \int_{\mathbf{R}} \frac{x^2 1_{|x| \leq Cn}}{1+x^2}\, dx = \frac{2}{\pi}(Cn - \tan^{-1}(Cn))$$

and hence $S_{n,\leq}/n$ has mean $\log(C^2 n^2 + 1)/\pi$ and variance at most $\frac{2/\pi \cdot (Cn - \tan^{-1}(Cn))}{n}$. By Chebyshev's inequality, we thus have

$$\mathbf{P}(|S_{n,\leq}/n - \log(C^2 n^2 + 1)/\pi| \geq \lambda) \leq \frac{2/\pi \cdot (Cn - \tan^{-1}(Cn))}{n\lambda^2}$$

for any $\lambda > 0$.

We now turn to $S_{n,>}$. Observe that the random variable $|X| 1_{|X|>Cn}$ is only nonzero with probability

$$1 - \mathbf{P}(|X| \leq Cn) = 1 - \frac{1}{\pi} \int_{-Cn}^{Cn} \frac{1}{1+x^2}\, dx = 1 - \frac{2\tan^{-1}(Cn)}{\pi}.$$

Thus $S_{n,>}$ is nonzero with probability at most $n(1 - \frac{2\tan^{-1}(Cn)}{\pi})$. By the triangle inequality, we conclude that

$$\mathbf{P}(|S_n/n - \log(C^2 n^2 + 1)/\pi| \geq \lambda) \leq \frac{2/\pi(Cn - \tan^{-1}(Cn))}{n\lambda^2} + n(1 - \frac{2\tan^{-1}(Cn)}{\pi})$$

which is valid for any $\lambda > 0$. If we pick $C = \lambda = \sqrt{\log n}$, we see that

$$\mathbf{P}(|S_n/n\log n - \log(n^2 \log n + 1)/\pi \log n| \geq \frac{1}{\sqrt{\log n}}) \leq O(\frac{1}{\log^{1/2} n})$$

which for large n implies that

$$\mathbf{P}(|S_n/n\log n - 2/\pi| \geq \frac{1}{\sqrt{\log n}}) \leq O(\frac{1}{\log^{1/2} n}),$$ giving the result.

Finally, to show that $\frac{|X_1|+\cdots+|X_n|}{n\log n}$ is almost surely unbounded, it suffices to show that $\frac{|X_n|}{n\log n}$ is.

Let $b_n = n\log n$. Note that $\forall x > 0, \mathbf{P}(|X| > x) = 1 - \frac{2\tan^{-1}(x)}{\pi} \sim \frac{2}{\pi x}$. In particular, we have $\mathbf{P}(|X| > x) \geq \frac{1}{2x}$ for sufficiently large x. Hence for any large M, it holds that $\sum_{n=2}^{\infty}\mathbf{P}(\frac{|X_n|}{bn} > M) \geq \sum_{n=2}^{\infty}\frac{1}{2Mb_n} = \infty$, and the claim follows from the second Borel-Cantelli lemma. □

Exercise 21

Proof. We form the partial row sum $S_n := X_{1,n,} + \cdots + X_{n,n}$ and the truncated row sums $S_{n,\leq} := X_{1,n,\leq} + \cdots + X_{n,n,\leq}, S_{n,>} := X_{1,n,>} + \cdots + X_{n,n,>}$. The aim is to show that $\forall \varepsilon > 0, \mathbf{P}(|\frac{S_n-\mu_n}{M_n}| > \varepsilon) \to 0$ as $n \to \infty$.

By the triangle inequality, we can split

$$\mathbf{P}(|\frac{S_n - \mu_n}{M_n}| > \varepsilon) \leq \mathbf{P}(|\frac{S_{n,\leq} - \mu_n}{M_n}| > \varepsilon) + \mathbf{P}(S_{n,>} \neq 0).$$

By Chebyshev's inequality, the first term on the RHS is less than

$$\frac{1}{\varepsilon^2 M_n^2}\mathbf{Var}(\mathbf{S_{n,\leq}}) \leq \frac{1}{\varepsilon^2 M_n^2}\sum_{i=1}^{n}\mathbf{E}|X_{i,n,\leq}|^2 \to 0 \text{ as } n \to \infty.$$

Also by triangle inequality again the second term on the RHS is less than

$\sum_{i=1}^{n}\mathbf{P}(|X_{i,n}| > M_n) \to 0$ as $n \to \infty$, hence we have

$\mathbf{P}(|\frac{S_n - \mu_n}{M_n}| > \varepsilon) \to 0$ as $n \to \infty$, completing the proof. □

Note on Lemma 22:

1. Intuitively, for sufficiently large n (and hence j), $n_{j,\delta} \approx (1+\delta)^j$ in the sense that the relative error $|\frac{(1+\delta)^j - n_{j,\delta}}{(1+\delta)^j}|$ can be made arbitrarily small. Thus one has $(1+2\delta)n_{j,\delta} \geq n_{j+1,\delta}$ and $\frac{1}{1+2\delta}n_{j+1,\delta} \leq n_{j,\delta}$ for large n, with the factor 2 serving as a safety margin here.

Precisely, by definition one has (for sufficiently large n)

$$(1+2\delta)n_{j,\delta} \geq (1+2\delta)((1+\delta)^j - 1) = (1+\delta)^{j+1} + \delta(1+\delta)^j - (1+2\delta) \geq n_{j+1,\delta}$$

from which it also follows that $\dfrac{1}{1+2\delta} n_{j+1,\delta} \leq n_{j,\delta}$.

2. We want to conclude that $S_n/n \xrightarrow{n} \mu$ holds almost surely, which by Lemma 22 is equivalent to showing that

$$\bigwedge_{\delta \to 0} \left(\frac{1}{1+2\delta}\mu \leq \liminf_{n\to\infty} \frac{S_n}{n} \leq \limsup_{n\to\infty} \frac{S_n}{n} \leq (1+2\delta)\mu\right).$$

holds almost surely. The countable intersection of events that occur almost surely, will continue to hold almost surely; but this is not necessarily the case when there are an uncountable number of events to be intersected. This is why it is necessary to place δ inside a countable set.

3. The mean and variance of the truncated average only requires pairwise independence to calculate; joint independence is overkill. (In general, k^{th} moments can be calculated assuming k-wise independence.) e.g. If the $X_1, \dots, X_n$ are pairwise independent in addition to being square-integrable, then all the covariances vanish, and we obtain additivity of the variance: $\mathbf{Var}(S_n) = \mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n)$.

Exercise 23

Proof. (1). By condition, the indicator random variables $1_{X_i \leq t}$ and $1_{X_i < t}$ are iid copies of the indicator random variable $1_{X \leq t}$ and $1_{X<t}$ respectively. Also, $|\{1 \leq i \leq n : X_i \leq t\}| = \sum_{i=1}^n 1_{X_i \leq t}, |\{1 \leq i \leq n : X_i < t\}| = \sum_{i=1}^n 1_{X_i < t}$, the conclusion then follows from the strong law of large numbers.

(2). Let $F_n(t) := \frac{1}{n}|\{1 \leq i \leq n : X_i \leq t\}|$ denote the empirical distribution function and $F(t) := \mathbf{P}(X \leq t)$ the CDF. For any natural number m, let $f_m(x)$ denote the largest integer multiple of $1/m$ less than x.

By construction, f_m is a piecewise constant function on the unit interval, divided at $0, 1/m, 2/m, \dots, 1$ (we let $f_m(0) := 0$). Subsequently $\mathbf{R}$ is partitioned into m intervals, divided at $t_0 = F^{-1}(0), t_1 = F^{-1}(1/m), \dots, t_m = F^{-1}(1)$, here we let $t_0 = F^{-1}(0) := -\infty$ and $t_m = F^{-1}(1) := \infty$.

For any $t \in \mathbf{R}$, $t \in [t_{j-1}, t_j]$ for some j, and correspondingly $F(t) \in [\frac{j-1}{m}, \frac{j}{m}]$. Since all of f_m, F, F_n are non-decreasing, we have

$$\begin{aligned} f_m(F_n(t)) - f_m(F(t)) &\leq f_m(F_n(t_j)) - f_m(F(t_{j-1})) \\ &= f_m(F_n(t_j)) - f_m(F(t_j)) + 1/m. \end{aligned}$$

Therefore,

$$\begin{aligned} \|f_m(F_n) - f_m(F)\|_\infty &= \sup_{t \in \mathbf{R}} |f_m(F_n(t)) - f_m(F(t))| \\ &\leq \max_{j \in \{1,\dots,m\}} |f_m(F_n(t_j)) - f_m(F(t_j))| + 1/m. \end{aligned}$$

By the strong law of large numbers (and the definition of f_m), we see that $\forall t \in \mathbf{R}, |f_m(F_n(t)) - f_m(F(t))| = O(1/m)$ almost surely for sufficiently large n. By the triangle inequality, we thus have

$$\begin{aligned} |F_n(t) - F(t)| &\leq |F_n(t) - f_m(F_n(t))| + |F(t) - f_m(F(t))| + |f_m(F_n(t)) - f_m(F(t))| \\ &= O(1/m) \end{aligned}$$

almost surely for sufficiently large n, as $1/m$ can be arbitrarily small, this proves the given claim. □

Exercise 24

Proof. (1). $\mathbf{E}|X| = \mathbf{E}X = \sum_{n=1}^{\infty} n\mathbf{P}(X = n) = \sum_{n=1}^{\infty} \frac{1}{\zeta(3)} \frac{1}{n^2} = \frac{\zeta(2)}{\zeta(3)} < \infty$, i.e. X is absolutely-integrable.

(2). For any constant $A > 0$, consider the events $X_{i,n} > nA$ for $i = 1, \dots, n$. By definition, we have $\mathbf{P}(X_{i,n} > nA) = \sum_{j \geq \lceil nA \rceil} \frac{1}{\zeta(3)} \frac{1}{j^3} \geq \frac{1}{\zeta(3)} \int_{\lceil nA \rceil}^{\infty} \frac{1}{x^3}\ dx \geq \frac{1}{\zeta(3)} \frac{1}{2(nA+1)^2} \geq \frac{\varepsilon}{n^2}$, for some $\varepsilon > 0$ depending on A.

Denote the sum of row n by S_n, it follows that

$$\mathbf{P}(S_n > nA) \geq \mathbf{P}(\bigvee_{i=1}^{n} (X_{i,n} > nA)) = 1 - \mathbf{P}(\bigwedge_{i=1}^{n} (X_{i,n} \leq nA)) \geq 1 - (1 - \frac{\varepsilon}{n^2})^n.$$

Using binomial approximation, $(1 - \frac{\varepsilon}{n^2})^n \approx 1 - \frac{\varepsilon}{n}$ (with the error term absorbed into the ε term), and we have $\mathbf{P}(S_n > nA) \geq \frac{\varepsilon}{n}$ for some $\varepsilon = \varepsilon(A)$. As the ε/n are not summable in n, the second Borel-Cantelli lemma implies that almost surely an infinite number of the events $S_n/n > A$ holds simultaneously, i.e. the row averages are almost surely unbounded. □

Exercise 28

Proof. Write $T_n := Y_1 + \cdots + Y_n$, and $\lim_{n\to\infty} T_n := s$. By the summation by parts formula, we have $\sum_{i=1}^n b_i Y_i = b_n T_n + \sum_{i=0}^{n-1} T_i(b_i - b_{i+1})$. For a fixed $\varepsilon > 0$, we choose $N > 0$ sufficiently large such that $|T_i - s| < \varepsilon$ for all $i \geq N$. It follows that

$$\begin{aligned}\frac{1}{b_n}\sum_{i=1}^n b_i Y_i &= T_n + \sum_{i=0}^{N-1} \frac{1}{b_n} T_i(b_i - b_{i+1}) + \sum_{i=N}^{n-1} \frac{1}{b_n} T_i(b_i - b_{i+1}) \\ &= T_n + \frac{1}{b_n}\sum_{i=0}^{N-1} T_i(b_i - b_{i+1}) + \sum_{i=N}^{n-1} \frac{1}{b_n}(T_i - s)(b_i - b_{i+1}) + \sum_{i=N}^{n-1} \frac{1}{b_n} s(b_i - b_{i+1}) \\ &\leq T_n + \frac{C_\varepsilon}{b_n} + (\frac{b_N}{b_n} - 1)\varepsilon + s(\frac{b_N}{b_n} - 1),\end{aligned}$$

where C_ε is a constant depending on ε. Sending $n \to \infty$ then yields the given result. □

Exercise 29

Proof. Let $S_{n,\leq A} = X_{1,\leq A} + \cdots + X_{n,\leq A}$, where $X_{i,\leq A} := X_i 1_{|X_i| \leq A}$, and $\mathbf{E} X_{i,\leq A} = \mu_{i,\leq A}$. By Theorem 26, the series $\sum_{i=1}^\infty (X_{i,\leq A} - \mu_{i,\leq A}) = \sum_{i=1}^\infty X_{i,\leq A} - \sum_{i=1}^\infty \mu_{i,\leq A}$ is almost surely convergent, with $\sum_{i=1}^\infty \mu_{i,\leq A} < \infty$ by condition. Hence the series $\sum_{i=1}^\infty X_{i,\leq A}$ is almost surely convergent.

As $X_i = X_{i,\leq A} + X_{i,>A}$, it remains to show that the series $\sum_{i=1}^\infty X_{i,>A}$ is almost surely convergent. By condition and the Borel-Cantelli lemma, this sum is almost surely a finite sum, i.e. almost surely convergent. □

Exercise 30

Proof. Let $Y_i := \dfrac{X_i}{i^{1/2}\log^{1/2+\varepsilon} i}$. By condition, we have

$$\sum_{i=1}^\infty \mathbf{Var}(Y_i) = \sum_{i=1}^\infty \frac{1}{i\log^{1+\varepsilon} i}\mathbf{Var}(X_i) \leq \sup_i \mathbf{E}(X_i) \sum_{i=1}^\infty \frac{1}{i\log^{1+\varepsilon} i} < \infty$$ (by the Cauchy criterion). Hence by Theorem 26, the series $\sum_{i=1}^\infty Y_i$ is almost surely convergent. By the Kronecker lemma, the sequence $\dfrac{S_n}{n^{1/2}\log^{1/2+\varepsilon} n}$ thus converges almost surely to zero as $n \to \infty$. □

Exercise 31

Proof. By the strong law of large numbers, the normalized sequence $Y_n := |S_n/n - \mu|$ converges almost surely to 0. By the triangle inequality and the absolute integrable condition, $\mathbf{E}|Y_n|^{1+\varepsilon} < \infty$ for any $\varepsilon > 0$. Thus by Theorem 25 of Note 1 (convergence for random variables with bounded moment), $\lim_{n\to\infty} \mathbf{E}|S_n/n - \mu| = 0$.

Alternatively, let $M > 0$. By dominated convergence $\mathbf{E}|X|1_{|X|\leq M} \to \mathbf{E}|X|$ as $M \to \infty$, so correspondingly $\mathbf{E}|X|1_{|X|>M} \to 0$ as $M \to \infty$. This implies that X_n is uniformly integrable by Exercise 1.5.11 of "M". The conclusion then follows from Theorem 1.5.13 of "M" and the weak law of large numbers. □

Exercise 32

Proof. Let $\mu := \mathbf{E}X$. When X is absolutely integrable, the weak law of large numbers gives $\mathbf{P}(|\frac{S_n}{\mu n} - 1| > \varepsilon) \to 0$ as $n \to \infty$, so one can simply set $a_n := \mu n$.

Now, suppose that X is not integrable. The definition of weak L^1 random variables implies fast shrinking tails, hence we use a truncation argument to split $X_i = X_{i,\leq M} + X_{i,>M}$, where $X_{i,\leq M} := X_i 1_{X_i \leq M}$ and $X_{i,>M} := X_i 1_{X_i > M}$, and then similarly decompose $S_n = S_{n,\leq} + S_{n,>}$ where $S_{n,\leq} := X_{1,\leq M} + \cdots + X_{n,\leq M}$ and $S_{n,>} := X_{1,>M} + \cdots + X_{n,>M}$. Denote by F the distribution function of X.

The random variable $X1_{X\leq M}$ has mean $\mathbf{E}(X1_{X\leq M}) = \mu(M)$ (which clearly tends to infinity as $M \to \infty$), and variance

$$\begin{aligned}\mathbf{Var}(X1_{X\leq M}) \leq \mathbf{E}X^2 1_{X\leq M} &= \int_0^M 2x(1-F(x))\,dx = \int_0^M 2x\mathbf{P}(X > x)\,dx \\ &\leq 2CM,\end{aligned}$$

where $C = \sup_{t>0} t\mathbf{P}(X \geq t)$. By Chebyshev's inequality, we thus have

$$\mathbf{P}(|\frac{S_{n,\leq}}{n} - \mu(M)| \geq \lambda) \leq \frac{2MC}{n\lambda^2} \text{ for any } \lambda > 0.$$

By the the Glivenko-Cantelli theorem: almost surely, one has

$$\frac{1}{n}|\{1 \leq i \leq n : X_i \leq t\}| \to \mathbf{P}(X \leq t) \text{ uniformly in } t \text{ as } n \to \infty.$$

Meanwhile, $\mathbf{P}(X \leq M) = 1 - \mathbf{P}(X > M) \geq 1 - C/M$. So almost surely, for sufficiently large n, one has $|\{1 \leq i \leq n : X_i \leq M\}| \geq n(1 - C/M)$, or equivalently $|\{1 \leq i \leq n : X_i > M\}| \leq nC/M$. This implies that almost surely, for sufficiently large n, $S_{n,>}$ is nonzero with probability at most nC^2/M^2. We conclude that almost surely,

$$\mathbf{P}(|\frac{S_n}{n} - \mu(M)| \geq \lambda) \leq \frac{2MC}{n\lambda^2} + \frac{nC^2}{M^2}, \text{ valid for any } \lambda, M > 0 \text{ and large } n.$$

Set $M = n$ and $\lambda = \varepsilon\mu(M)$, we get almost surely,

$$\mathbf{P}(|\frac{S_n}{n\mu(n)} - 1| \geq \varepsilon) \leq \frac{2C}{\varepsilon^2\mu^2(n)} + C^2/n, \text{ valid for large } n.$$

Sending $n \to \infty$, the RHS tends to zero, the claim stands with $a_n = n\mu(n)$. $\square$

Remark on Exercise 32:

If one controls $\mathbf{P}(S_{n,>} \neq 0)$ instead by $n\mathbf{P}(X > M)$, with $M = M(n) = b_n$ and $\lambda = \varepsilon\mu(b_n)$, one ends up with the bound

$$\mathbf{P}(|\frac{S_n}{n\mu(b_n)} - 1| \geq \varepsilon) \leq \frac{2b_nC}{n\varepsilon^2\mu^2(b_n)} + nO(\frac{1}{b_n}).$$

One therefore aims to choose the truncation parameter $M = b_n$ obeying $(1).b_n = o(n\mu^2(b_n))$ and $(2).n = o(b_n)$. Per condition (2), we tentatively set $b_n = nr(n)$ for some $r(n) \to \infty$. If we pick $r(n)$ with $\mu(nr(n))/r(n) \geq 1$ for large n, then in particular $b_n \leq n\mu(b_n)$ and condition (1) is met. For instance, consider the sequence $(n_m)_m$ defined inductively as

$$n_m := \inf\{k \geq n_{m-1} : \forall_{j \geq k}, \mu(j) \geq 2^m\}, n_0 := 0.$$

Let r be an increasing step function constant on $n_{m-1}+1, n_{m-1}, \ldots, n_m$ with $1 \leq r(n_m) \leq 2^m$ for all m. Then $n_m r(n_m) \geq n_m$ and $\mu(n_m r(n_m)) \geq 2^m$, giving $\mu(n_m r(n_m))/r(n_m) \geq 1$ for all m. The result thus holds with $a_n = n\mu(b_n)$. Refer to Theorem 2, p.236 of “An Introduction to Probability Theory and Its Applications”, vol 2.

Exercise 33

Proof. If X is symmetric, then from

$$\mathbf{P}(X \geq 0) = \mathbf{P}(-X \geq 0) = \mathbf{P}(X \leq 0) = \mathbf{P}(X < 0) + \mathbf{P}(X = 0)$$

one gets $2\mathbf{P}(X \geq 0) = 1 + \mathbf{P}(X = 0)$, so $\mathbf{P}(X \geq 0) \geq 1/2$. Moreover, if X and Y are both real symmetric random variables, then

$$\begin{aligned}\mathbf{P}(X + Y \leq t) &= \mathbf{P}(X \leq t - Y) = \mathbf{P}(-X \leq t - Y) = \mathbf{P}(-X \leq t + Y) \\ &= \mathbf{P}(-X - Y \leq t),\end{aligned}$$

and in particular the sum $X + Y$ is also a real symmetric random variable. By induction, the sum of any finite number of real symmetric random variables is again symmetric, we thus have $\mathbf{P}(S_n - S_i) = \mathbf{P}(X_{i+1} + \cdots + X_n) \geq 1/2$ for each $1 \leq i \leq n$.

For each i, let E_i be the event that $S_i \geq \lambda$, but that $S_j < \lambda$ for all $1 \leq j < i$. It is clear that the event $\sup_{1 \leq i \leq n} S_i \geq \lambda$ is the disjunction of the disjoint events $E_1, \ldots, E_n$, thus

$$\mathbf{P}(\sup_{1 \leq i \leq n} S_i \geq \lambda) = \sum_{i=1}^{n} \mathbf{P}(E_i).$$

On the other hand, we have the inclusion $(S_n \geq \lambda) \supset \bigvee_{i=1}^{n} (E_i \wedge (S_n - S_i \geq 0))$, thus by independence of the S_i and $S_n - S_i$ we get

$$\begin{aligned}\mathbf{P}(S_n \geq \lambda) \geq \mathbf{P}(\bigvee_{i=1}^{n} (E_i \wedge (S_n - S_i \geq 0))) &= \sum_{i=1}^{n} \mathbf{P}(E_i)\mathbf{P}(S_n - S_i \geq 0) \\ \geq \frac{1}{2} \sum_{i=1}^{n} \mathbf{P}(E_i) &= \frac{1}{2}\mathbf{P}(\sup_{1 \leq i \leq n} S_i \geq \lambda)\end{aligned}$$

, and we are done. □

Exercise 34

Proof. From the inclusion

$$(S_n - S_i < -\lambda) \subset (|S_n| - |S_i| < -\lambda) = (|S_i| > \lambda + |S_n|) \subset (|S_i| \geq \lambda),$$

it follows that $\mathbf{P}(S_n - S_i < -\lambda) < \varepsilon$, or equivalently that $\mathbf{P}(S_n - S_i \geq -\lambda) \geq 1 - \varepsilon$. For each i, let E_i be the event that $S_i \geq 2\lambda$, but that $S_j < 2\lambda$ for all $1 \leq j < i$. It is clear that the event $\sup_{1 \leq i \leq n} S_i \geq 2\lambda$ is the disjunction of the disjoint events $E_1, \ldots, E_n$, thus

$$\mathbf{P}(\sup_{1 \leq i \leq n} S_i \geq 2\lambda) = \sum_{i=1}^{n} \mathbf{P}(E_i).$$

Note that $E_i \wedge (S_n - S_i \geq -\lambda) \subset (S_n \geq \lambda) \wedge E_i$, which implies that

$$\sum_{i=1}^{n} \mathbf{P}(E_i \wedge (S_n - S_i \geq -\lambda)) \leq \mathbf{P}(S_n \geq \lambda) \leq \mathbf{P}(|S_n| \geq \lambda) \leq \varepsilon$$

By independence of the S_i and $S_n - S_i$,

$$\mathbf{P}(E_i \wedge (S_n - S_i \geq -\lambda)) = \mathbf{P}(E_i)\mathbf{P}(S_n - S_i \geq -\lambda) \geq (1-\varepsilon)\mathbf{P}(E_i).$$

Putting all this together, we obtain

$$(1-\varepsilon)\sum_{i=1}^{n} \mathbf{P}(E_i) = (1-\varepsilon)\mathbf{P}(\sup_{1\leq i\leq n} S_i \geq 2\lambda) \leq \varepsilon, \text{ or } \mathbf{P}(\sup_{1\leq i\leq n} S_i \geq 2\lambda) \leq \frac{\varepsilon}{1-\varepsilon},$$

as desired. $\square$