An Analytical Introduction to Probability Theory

7. Strong Law of Large Numbers

DASN, NTU https://personal.ntu.edu.sg/wptay/

7.1 SLLN

In this note, we prove the Strong Law of Large Numbers (SLLN). We give three different proofs of SLLN, which are based on the ideas of maximal inequality, truncation, and positivity. First, let us state the SLLN.

Theorem 7.1 (Kolmogorov's SLLN). *Suppose $X_1, X_2, \ldots$ are i.i.d. random variables, $\mathbb{E}|X_1| < \infty$, $\mathbb{E}X_1 = 0$, and $S_n = \sum_{i=1}^n X_i$. We have*

$$\frac{S_n}{n} \to 0 \ \text{a.s.} \ \text{as } n \to \infty.$$

7.2 The First Proof: L^1 Maximal Inequality

For a sequence $a_1, a_2, \ldots, a_N$, we say that a_j is an n-leader if $a_j + a_{j+1} + \ldots + a_k > 0$ for some k such that $j \le k \le N$ and $k - j + 1 \le n \le N$.

Lemma 7.1. *Let $G \neq \emptyset$ be the set of indices of all n-leaders of $a_1, \ldots, a_N$. Then*

$$\sum_{j \in G} a_j > 0.$$

Proof. Let

$$\begin{aligned} j_1 &= \min G, \\ k_1 &= \min\{k \ge j_1 : a_{j_1} + \ldots + a_k > 0\}. \end{aligned}$$

Since $j_1 \in G$, $k_1 \le N$ exists and $k_1 - j_1 + 1 \le n$. Then we have $\sum_{j=j_1}^{s-1} a_j < 0$ for $j_1 \le s \le k_1$. Therefore,

$$\sum_{k=s}^{k_1} a_j = \sum_{j=j_1}^{k_1} a_j - \sum_{j=j_1}^{s-1} a_j > 0,$$

This implies that $s \in G$ and hence $[j_1, k_1] \subset G$. Let $G_1 = G \setminus [j_1, k_1] \neq \emptyset$ and repeat the same process until $G_{l+1} = \emptyset$ for some l. We obtain

$$\bigcup_{i=1}^{l} [j_i, k_i] = G,$$

where $[j_i, k_i] \cap [j_m, k_m] = \emptyset, \ \forall\, i \neq m$. Finally, we have

$$\sum_{j \in G} a_j = \sum_{i=1}^{l} \sum_{s=j_i}^{k_i} a_s > 0.$$

□

Theorem 7.2 (L^1 maximal inequality)**.** *Let X_k, $k \geq 1$, be i.i.d. random variables. Then,* $\forall\, \epsilon > 0$,

$$\mathbb{P}\left(\max_{1 \leq k \leq n} \frac{S_k}{k} > \epsilon \right) \leq \frac{1}{\epsilon} \mathbb{E} X_1^+.$$

Proof. Fix $\epsilon > 0$ and $N > 1$. For $1 \leq n \leq N$, let

$$A_j = \left\{ \omega \,:\, \max_{\substack{1 \leq k \leq n \\ j+k-1 \leq N}} \frac{S(j, j+k-1)}{k} > \epsilon \right\},$$

where $S(j,k) = X_j + \ldots + X_k$. From the definition, we have

$$A_j = \{\omega \,:\, a_j(\omega) \text{ is } n\text{-leader in} (a_1, a_2, \ldots, a_N)\},$$

where $a_j(\omega) = X_j(\omega) - \epsilon$. Applying Lemma 7.1, we obtain

$$\sum_{j=1}^{N} (X_j - \epsilon) \mathbf{1}_{A_j} > 0,$$

which yields

$$\epsilon \sum_{j=1}^{N-n} \mathbf{1}_{A_j} \leq \epsilon \sum_{j=1}^{N} \mathbf{1}_{A_j} < \sum_{j=1}^{N} X_j \mathbf{1}_{A_j} \leq \sum_{j=1}^{N} X_j^+. \tag{1}$$

Furthermore, for $1 \leq j \leq N - n$, since the X_i's are i.i.d., we have

$$\mathbb{P}(A_j) = \mathbb{P}(A_1) = \mathbb{P}\left(\max_{1 \leq k \leq n} \frac{S_k}{k} > \epsilon \right).$$

Taking expectations in (1), we obtain

$$\begin{aligned} \epsilon (N - n) \mathbb{P}(A_1) &\leq N \mathbb{E} X_1^+ \\ \frac{N-n}{N} \mathbb{P}(A_1) &\leq \frac{1}{\epsilon} \mathbb{E} X_1^+. \end{aligned}$$

By taking $N \to \infty$, the proof is complete. □

Corollary 7.1. *Let $X_k, k \geq 1$ be i.i.d random variables. Then, $\forall\ \epsilon > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\frac{S_k}{k}\right| > \epsilon\right) \leq \frac{1}{\epsilon}\mathbb{E}|X_1|.$$

Proof. Apply the L^1 maximal inequality to (X_k) and $(-X_k)$. $\square$

We are now ready to give the first proof of the SLLN. Fix $\lambda > 0$. Let

$$X_k' = X_k \mathbf{1}_{\{|X_k|\leq\lambda\}},$$
$$X_k'' = X_k \mathbf{1}_{\{|X_k|>\lambda\}}.$$

As the threshold λ is fixed for all k, this is known as *fixed truncation.* We let

$$S_n' = X_1' + \ldots + X_n',$$
$$S_n'' = X_1'' + \ldots + X_n'',$$

so that $S_n = S_n' + S_n''$. We have

$$\begin{aligned}
&\mathbb{E}X_k' + \mathbb{E}X_k'' = \mathbb{E}X_k = 0,\\
&S_n = S_n' - n\mathbb{E}X_1' + S_n'' - n\mathbb{E}X_1'',\\
&\left|\frac{S_n}{n}\right| \leq \left|\frac{S_n'}{n} - \mathbb{E}X_1\mathbf{1}_{\{|X_1|\leq\lambda\}}\right| + \left|\frac{S_n''}{n} - \mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}}\right|.
\end{aligned}$$

Since X_k' has bounded moments, we can apply the SLLN with 2nd moments (Lemma 6.7) to obtain

$$\left|\frac{S_n'}{n} - \mathbb{E}X_1\mathbf{1}_{\{|X_1|\leq\lambda\}}\right| \to 0 \text{ a.s.}$$

Furthermore, we have

$$\begin{aligned}
L = \limsup_{n\to\infty}\left|\frac{S_n}{n}\right| &\leq \limsup_{n\to\infty}\left|\frac{S_n''}{n} - \mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}}\right|\\
&\leq \sup_{n\geq 1}\left|\frac{S_n''}{n} - \mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}}\right|\\
&\leq \sup_{n\geq 1}\left|\frac{S_n''}{n}\right| + \mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}}.
\end{aligned}$$

Since $\mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}} \leq \mathbb{E}|X_1| < \infty$, we apply DCT (Theorem 4.3) to obtain

$$\lim_{\lambda\to\infty}\mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}} = \mathbb{E}X_1\mathbf{1}_{\{|X_1|=\infty\}} = 0.$$

Therefore, $\forall\,\epsilon > 0$, $\exists\,\lambda_0$ such that

$$\mathbb{E}X_1\mathbf{1}_{\{|X_1|>\lambda\}} < \epsilon,\ \forall\,\lambda \geq \lambda_0.$$

By using the above result and applying Corollary 7.1, we obtain for all $\lambda \geq \lambda_0$,

$$\begin{aligned}\mathbb{P}(L > 2\epsilon) &\leq \mathbb{P}\left(\sup_{n\geq 1}\left|\frac{S_n''}{n}\right| > \epsilon\right) \\ &\leq \lim_{\lambda\to\infty}\frac{1}{\epsilon}\mathbb{E}|X_1|\mathbf{1}_{\{|X_1|>\lambda\}} = 0.\end{aligned}$$

Taking $\epsilon \to 0$, we have

$$\mathbb{P}(L > 0) = 0 \implies \mathbb{P}(L = 0) = 1 \implies \frac{S_n}{n} \to 0 \text{ a.s. as } n \to \infty.$$

The first proof of SLLN is now complete.

7.3 The Second Proof: Kolmogorov

We next discuss Kolmogorov's proof of SLLN, which was done around 1930.

Theorem 7.3 (Kolmogorov's maximal inequality)**.** *Suppose $X_k, k \geq 1$ are independent random variables and $\mathbb{E}|X_k| < \infty$, then*

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}\operatorname{var}(S_n),\ \forall\ \lambda > 0.$$

Note that $\operatorname{var}(S_n)$ always exists. If it is infinite, then the bound is trivial.

Proof. Without loss of generality, we assume $\mathbb{E}X_k = 0$. Let

$$\tau = \min\{1 \leq k \leq n\,:\,|S_k| \geq \lambda\},$$

where $\tau = \infty$ if $|S_k| < \lambda$ for all $1 \leq k \leq n$. Define

$$S_\tau = \sum_{k=1}^{n} S_k\mathbf{1}_{\{\tau=k\}}.$$

Since

$$\lambda^2\mathbf{1}_{\{\max_{1\leq k\leq n}|S_k|\geq\lambda\}} \leq S_\tau^2,$$

we obtain

$$\lambda^2\mathbb{P}\left(\max_{1\leq k\leq n}|S_k| \geq \lambda\right) \leq \mathbb{E}S_\tau^2.$$

It suffices to prove $\mathbb{E}S_\tau^2 \leq \mathbb{E}S_n^2$. We have

$$\begin{aligned}S_n &= S_\tau + S_n - S_\tau,\\ S_n^2 &= S_\tau^2 + (S_n - S_\tau)^2 + 2S_\tau\,(S_n - S_\tau)\,,\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{E} S_\tau \left(S_n - S_\tau\right) &= \mathbb{E}\left[\sum_{k=1}^{n} \mathbf{1}_{\{\tau=k\}} S_k \left(S_n - S_k\right)\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\big[\mathbf{1}_{\{\tau=k\}} S_k\big] \mathbb{E}\left(S_n - S_k\right) = 0,
\end{aligned}
$$

where the penultimate equality follows from independence. We thus have

$$
\mathbb{E} S_\tau^2 \le \mathbb{E} S_n^2,
$$

and the proof is complete. $\square$

Theorem 7.4 (Variance convergence criterion)**.** *Suppose $Y_k,\ k \ge 1$ are independent random variables and $\mathbb{E} Y_k = 0$. If $\sum_{k=1}^{\infty} \operatorname{var}(Y_k) < \infty$, then*

$$
\sum_{k=1}^{\infty} Y_k \textit{ converges a.s.}
$$

Proof. Let $S_n(\omega) = \sum_{k=1}^{n} Y_k(\omega)$. It suffices to show that $(S_n)_{n\ge 1}$ is Cauchy a.s., i.e.,

$$
R_N = \sup_{n,m \ge N} |S_n - S_m| \to 0 \text{ a.s. as } N \to \infty.
$$

Let $N \ge N_0 \ge 1$. We have

$$
\begin{aligned}
R_N &\le R_{N_0} \\
&\le \sup_{n \ge N_0} |S_n - S_{N_0}| + \sup_{m \ge N_0} |S_m - S_{N_0}|.
\end{aligned}
$$

Therefore, $\forall\, \epsilon > 0$,

$$
\begin{aligned}
\mathbb{P}\left(\limsup_{N\to\infty} R_N > \epsilon\right) &\le 2\mathbb{P}\left(\sup_{n \ge N_0} |S_n - S_{N_0}| > \frac{\epsilon}{2}\right) \\
&\le \frac{8}{\epsilon^2} \sum_{k=N_0+1}^{\infty} \operatorname{var}(Y_k),
\end{aligned}
$$

where the last inequality follows from Chebyshev's inequality. Since $\sum_{k=1}^{\infty} \operatorname{var}(Y_k) < \infty$, we have

$$
\begin{aligned}
&\lim_{N_0\to\infty} \sum_{k=N_0+1}^{\infty} \operatorname{var}(Y_k) = 0 \\
\implies &\mathbb{P}\left(\limsup_{N\to\infty} R_N > \epsilon\right) = 0.
\end{aligned}
$$

$\square$

Let $\hat{X}_k = X_k \mathbf{1}_{\{|X_k| \le k\}}$. This is called *moving truncation*. We have

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{P}\Big(\hat{X}_k \neq X_k\Big) &= \sum_{k\ge 1} \mathbb{P}(|X_k| > k)\\ &= \sum_{k\ge 1} \mathbb{P}(|X_1| > k)\\ &\le \mathbb{E}|X_1| < \infty.\end{aligned}$$

From the first Borel-Cantelli lemma (Lemma 6.1), we obtain $\mathbb{P}\Big(\hat{X}_k \neq X_k \text{ i.o}\Big) = 0$, which means that for a.s. all $\omega \in \Omega$, $\exists\, K(\omega)$ s. t. $X_k(\omega) = \hat{X}_k(\omega), \forall\, k \ge K(\omega)$. Therefore, $\frac{1}{n}\sum_{k=1}^n X_k \to 0$ a.s. if and only if $\frac{1}{n}\sum_{k=1}^n \hat{X}_k \to 0$ a.s. In the rest of this section, we prove $\frac{1}{n}\sum_{k=1}^n \hat{X}_k \to 0$ a.s..

Lemma 7.2. *Suppose X_k, $k \ge 1$ are identically distributed random variables and $\hat{X}_k = X_k \mathbf{1}_{\{|X_k| \le k\}}$. We have*

$$\sum_{k=1}^{\infty} \frac{\operatorname{var}(\hat{X}_k)}{k^2} \le 2\mathbb{E}|X_1|.$$

Proof.

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{\operatorname{var}(\hat{X}_k)}{k^2} &\le \sum_{k=1}^{\infty} \frac{\mathbb{E}\hat{X}_k^2}{k^2}\\ &= \sum_{k=1}^{\infty} \frac{1}{k^2}\mathbb{E}\big[X_1^2 \mathbf{1}_{\{|X_1| \le k\}}\big]\\ &= \mathbb{E}\left[X_1^2 \sum_{k=1}^{\infty} \frac{1}{k^2}\mathbf{1}_{\{|X_1| \le k\}}\right]\\ &\le 2\mathbb{E}|X_1|,\end{aligned}$$

where the last equality comes from Fubini's theorem (Theorem 5.2) and the last inequality is because

$$\phi(x) = x^2 \sum_{k\ge 1} \frac{1}{k^2}\mathbf{1}_{\{x \le k\}} \le 2x$$

for all $x \ge 0$. To prove this, we note that

$$\int_{i-1}^{\infty} \frac{1}{x^2}\,\mathrm{d}x = \frac{1}{i-1} \ge \sum_{k=i}^{\infty} \frac{1}{k^2}.$$

If $0 \le x \le 1$, we have

$$\phi(x) = x^2 \sum_{k=1}^{\infty} \frac{1}{k^2} = x^2 \left(1 + \sum_{k=2}^{\infty} \frac{1}{k^2}\right) \le 2x^2 \le 2x.$$

A similar reasoning applies for $1 < x \le 2$ and $2 < x < \infty$ and the proof is complete. $\square$

Lemma 7.3 (Cesaro's Lemma)**.** *If* $\lim_{n\to\infty} a_n = a$, *then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = a.$$

Proof. Without loss of generality, we assume $a = 0$. For $k \le n$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n} a_i\right| \le \frac{1}{n}\left|\sum_{i=1}^{k} a_i\right| + \frac{n-k}{n} \sup_{j>k} |a_j|,$$
$$\Longrightarrow \limsup_{n\to\infty} \frac{1}{n}\left|\sum_{i=1}^{n} a_i\right| \le \limsup_{j\to\infty} |a_j| = 0.$$

$\square$

Lemma 7.4 (Kronecker's Lemma)**.** *If* $\sum_{k=1}^{\infty} \frac{a_k}{k}$ *converges, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = 0.$$

Proof. Let $r_n = \sum_{i=n+1}^{\infty} \frac{a_i}{i}$. We have

$$\lim_{n\to\infty} r_n = 0.$$

From Cesaro's lemma (Lemma 7.3), we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} r_k = 0.$$

Furthermore, for $k < n$, we have

$$r_k = \frac{a_{k+1}}{k+1} + \frac{a_{k+2}}{k+2} + \cdots + \frac{a_n}{n} + r_n.$$

Therefore,

$$\sum_{k=0}^{n-1} r_k = a_1 + \ldots + a_n + nr_n,$$
$$\implies \frac{1}{n}\sum_{k=1}^{n} a_k = -r_n + \frac{1}{n}\sum_{k=0}^{n-1} r_k.$$

By using the above results, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} a_k = 0.$$

□

We are now ready to give Kolmogorov's proof of the SLLN. From Lemma 7.2, we have

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\frac{1}{k^2}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right)^2\right] < \infty.$$

From Theorem 7.4, we obtain

$$\sum_{k=1}^{\infty} \frac{1}{k}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right) \text{ converges a.s.}$$

From Kronecker's lemma (Lemma 7.4), we have

$$\frac{1}{n}\sum_{k=1}^{n}\left(\hat{X}_k - \mathbb{E}\hat{X}_k\right) \to 0 \text{ a.s.} \tag{2}$$

Furthermore, using the DCT, we have

$$\lim_{k\to\infty} \mathbb{E}\hat{X}_k = \mathbb{E}\left[\lim_{k\to\infty} X_1 \mathbf{1}_{\{|X_1|\le k\}}\right] = \mathbb{E}X_1 = 0, \tag{3}$$
$$\implies \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\hat{X}_k \to 0 \text{ a.s..} \tag{4}$$

Therefore, from (2), we obtain

$$\frac{1}{n}\sum_{k=1}^{n} \hat{X}_k \to 0 \text{ a.s.,}$$

and the SLLN is proved.

We now give a generalization of Kronecker's lemma, which will be useful later on.

Lemma 7.5 (Generalized Kronecker's Lemma)**.** *Suppose* $0 < b_n \to \infty$ *as* $n \to \infty$*. If* $\sum_{n=1}^{\infty} \frac{a_n}{b_n}$ *converges in* $\mathbb{R}$*, then*

$$\frac{1}{b_n} \sum_{k=1}^{n} a_k \to 0 \text{ as } n \to \infty.$$

Proof. Without loss of generality, we can suppose that $b_n \in \mathbb{Z}_+$, $\forall\, n \geq 1$. Let $b_0 = 0$, $s_0 = 0$ and

$$s_n = \sum_{k=1}^{n} \frac{a_k}{b_k} \to s \in \mathbb{R}.$$

It can be checked that

$$\frac{1}{b_n} \sum_{k=1}^{n} a_k = s_n - \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) s_{k-1} \to 0,$$

since from Cesaro's lemma (note that $b_k - b_{k-1} \in \mathbb{Z}_+$ and $b_n = \sum_{k=1}^{n} (b_k - b_{k-1})$), we have

$$\lim_{n \to \infty} \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) s_{k-1} = s.$$

□

Theorem 7.5 (Generalized variance convergence criterion)**.** *Suppose* $Y_k, k \geq 1$ *are independent random variables,* $\mathbb{E}Y_k = 0$*, and* $0 < b_k \to \infty$ *as* $k \to \infty$*. We have*

$$\sum_{k=1}^{\infty} \frac{\operatorname{var}(Y_k)}{b_k^2} < \infty \implies \frac{1}{b_n} \sum_{k=1}^{\infty} Y_k \to 0 \text{ a.s.}$$

Proof. Apply Theorem 7.4 to Y_k / b_k and the result follows from Lemma 7.5. □

7.4 The Third Proof: Etemadi's Use of Positivity

In the third proof, we present the surprising and elegant proof of Etemadi discovered in 1981, some 50 years after Kolmogorov first proved the SLLN. This proof also strengthens the result to require only pairwise independence.

Theorem 7.6 (Etemadi's SLLN)**.** *Suppose* X_i*,* $i \geq 1$ *are identically distributed and pairwise independent, and* $\mathbb{E}X_i = \mu$*. We have*

$$\frac{S_n}{n} \to \mu \text{ a.s. as } n \to \infty.$$

Proof. We first observe that $X_i = X_i^+ - X_i^-$ and

$$X_i \perp\!\!\!\perp X_j \implies X_i^+ \perp\!\!\!\perp X_j^+,\; X_i^- \perp\!\!\!\perp X_j^-.$$

Therefore, we can without loss of generality assume $X_i \geq 0$. This turns out to be the key of Etemadi's proof. Let

$$\begin{aligned}\hat{X}_i &= X_i \mathbf{1}_{\{X_i \leq i\}},\\ \hat{S}_n &= \sum_{i=1}^{n} \hat{X}_i.\end{aligned}$$

As shown previously, it suffices to show that $\hat{S}_n/n \to \mu$ a.s. Let $\alpha \in (1,2)$, and $j_n = \lfloor \alpha^n \rfloor$. We have

$$\begin{aligned}1 \leq j_n \leq \alpha^n < j_{n+1} \leq 2j_n,\\ \implies \frac{1}{j_n} \leq \frac{2}{\alpha^n}.\end{aligned}$$

For a fixed i, let $n_0 = \min\{n : \alpha^n \geq i\}$. We have

$$\sum_{n:j_n \geq i} \frac{1}{j_n^2} \leq \sum_{n:j_n\geq i} \frac{4}{\alpha^{2n}} = 4\sum_{k=0}^{\infty} \frac{1}{\alpha^{2k}\alpha^{2n_0}} \leq 4\sum_{k=0}^{\infty} \frac{1}{i^2}\frac{1}{\alpha^{2k}} = \frac{4}{i^2}\frac{1}{1-\alpha^{-2}}. \tag{5}$$

For $\epsilon > 0$ and using Chebyshev's inequality together with pairwise independence, we have

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\hat{S}_{j_n} - \mathbb{E}\hat{S}_{j_n}\right| \geq \epsilon j_n\right) &\leq \frac{1}{\epsilon^2}\sum_{n=1}^{\infty}\frac{1}{j_n^2}\operatorname{var}(\hat{S}_{j_n})\\ &= \frac{1}{\epsilon^2}\sum_{n=1}^{\infty}\frac{1}{j_n^2}\sum_{i=1}^{j_n}\operatorname{var}(\hat{X}_i)\\ &= \frac{1}{\epsilon^2}\sum_{i=1}^{\infty}\operatorname{var}(\hat{X}_i)\sum_{n:j_n\geq i}\frac{1}{j_n^2}\\ &\leq \frac{4}{\epsilon^2}\frac{1}{1-\alpha^{-2}}\sum_{i=1}^{\infty}\frac{\operatorname{var}(\hat{X}_i)}{i^2}\\ &\leq \frac{8}{\epsilon^2}\frac{1}{1-\alpha^{-2}}\mathbb{E}|X_1| < \infty,\end{aligned}$$

where the penultimate inequality follows from (5), and the last equality from Lemma 7.2. From Lemma 6.5, we obtain

$$\frac{\hat{S}_{j_n} - \mathbb{E}\hat{S}_{j_n}}{j_n} \to 0 \text{ a.s.}$$

Using the same argument as in (4), we have

$$\mathbb{E}\hat{X}_i \to \mathbb{E}X_i = \mu \implies \frac{\mathbb{E}\hat{S}_{j_n}}{j_n} \to \mu.$$

Therefore, we obtain

$$\frac{\hat{S}_{j_n}}{j_n} \to \mu \text{ a.s.}$$

Finally, for any n, there exists k such that

$$j_k \le n < j_{k+1},$$

and because $\hat{X}_i \ge 0$ for all $i \ge 1$, we have a.s.,

$$\begin{aligned}
& \hat{S}_{j_k} \le \hat{S}_n \le \hat{S}_{j_{k+1}}, \\
\implies & \frac{j_k}{j_{k+1}} \frac{\hat{S}_{j_k}}{j_k} \le \frac{\hat{S}_n}{n} \le \frac{j_{k+1}}{j_k} \frac{\hat{S}_{j_{k+1}}}{j_{k+1}}, \\
\implies & \lim_{n\to\infty} \frac{j_k}{j_{k+1}} \frac{\hat{S}_{j_k}}{j_k} \le \liminf_{n\to\infty} \frac{\hat{S}_n}{n} \le \limsup_{n\to\infty} \frac{\hat{S}_n}{n} \le \lim_{n\to\infty} \frac{j_{k+1}}{j_k} \frac{\hat{S}_{j_{k+1}}}{j_{k+1}}, \\
\implies & \frac{1}{\alpha}\mu \le \liminf_{n\to\infty} \frac{\hat{S}_n}{n} \le \limsup_{n\to\infty} \frac{\hat{S}_n}{n} \le \alpha\mu.
\end{aligned}$$

Taking $\alpha \to 1$, we obtain

$$\lim_{n\to\infty} \frac{\hat{S}_n}{n} = \mu \text{ a.s.}$$

$\square$