9. Strong Law of Large Numbers

Our goal at this point is to strengthen the conclusion of Theorem 7.3 from convergence in probability to almost sure convergence. The following proof is due to Nasrollah Etemadi.

Theorem 9.1 (Strong Law of Large Numbers)**.** *Suppose that $X_1, X_2, \ldots$ are pairwise independent and identically distributed with $E|X_1| < \infty$. Let $S_n = \sum_{k=1}^n X_k$ and $\mu = E[X_1]$. Then $\frac{1}{n}S_n \to \mu$ almost surely as $n \to \infty$.*

Proof.
We begin by noting that $X_k^+ = \max\{X_k, 0\}$ and $X_k^- = \max\{-X_k, 0\}$ satisfy the theorem's assumptions, so, since $X_k = X_k^+ - X_k^-$, we may suppose without loss of generality that the $X_k's$ are nonnegative.

Next, we observe that it suffices to consider truncated versions of the $X_k's$:

Claim 9.1. If $Y_k = X_k 1\{X_k \le k\}$ and $T_n = \sum_{k=1}^n Y_k$, then $\dfrac{1}{n}T_n \to \mu$ a.s. implies $\dfrac{1}{n}S_n \to \mu$ a.s.

Proof. Lemma 7.2 and the fact that $G(t) = P(X_1 > t)$ is nonincreasing imply

$$\sum_{k=1}^{\infty} P(X_k \neq Y_k) = \sum_{k=1}^{\infty} P(X_k > k) = \sum_{k=1}^{\infty} P(X_1 > k) \le \int_0^{\infty} P(X_1 > t)\, dt = E|X_1| < \infty,$$

so the first Borel-Cantelli lemma gives $P(X_k \neq Y_k \text{ i.o.}) = 0$. Thus for all ω in a set of probability one, $\sup_n |S_n(\omega) - T_n(\omega)| < \infty$, hence $\dfrac{S_n}{n} - \dfrac{T_n}{n} \to 0$ a.s. and the claim follows. □

The truncation step should not be too surprising as it is generally easier to work with bounded random variables. The reason that we reduced the problem to the $X_k \ge 0$ case is that this assures that the sequence $T_1, T_2, \ldots$ is nondecreasing.

Our strategy will be to prove convergence along a cleverly chosen subsequence and then exploit monotonicity to handle intermediate values.

Specifically, for $\alpha > 1$, let $k(n) = \lfloor \alpha^n \rfloor$, the greatest integer less than or equal to α^n.
Chebychev's inequality and Tonelli's theorem give

$$\sum_{n=1}^{\infty} P\left(\left|T_{k(n)} - E\left[T_{k(n)}\right]\right| > \varepsilon k(n)\right) \le \sum_{n=1}^{\infty} \frac{\mathrm{Var}\left(T_{k(n)}\right)}{\varepsilon^2 k(n)^2} = \varepsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m)$$
$$= \varepsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n:k(n)\ge m} k(n)^{-2} \le \varepsilon^{-2} \sum_{m=1}^{\infty} E\left[Y_m^2\right] \sum_{n:\alpha^n \ge m} \lfloor \alpha^n \rfloor^{-2}.$$

Since $\lfloor \alpha^n \rfloor \ge \frac{1}{2}\alpha^n$ for $n \ge 1$ (by casing out according to α^n smaller or bigger than 2),

$$\sum_{n:\alpha^n \ge m} \lfloor \alpha^n \rfloor^{-2} \le 4 \sum_{n \ge \log_\alpha m} \alpha^{-2n} \le 4\alpha^{-2\log_\alpha m} \sum_{n=0}^{\infty} \alpha^{-2n} = 4(1-\alpha^{-2})^{-1} m^{-2},$$

hence

$$\sum_{n=1}^{\infty} P\left(\left|T_{k(n)} - E\left[T_{k(n)}\right]\right| > \varepsilon k(n)\right) \le \varepsilon^{-2} \sum_{m=1}^{\infty} E\left[Y_m^2\right] \sum_{n:\alpha^n \ge m} [\alpha^n]^{-2}$$
$$\le 4(1-\alpha^{-2})^{-1}\varepsilon^{-2} \sum_{m=1}^{\infty} \frac{E\left[Y_m^2\right]}{m^2}.$$

At this point, we note that

Claim 9.2. $\displaystyle\sum_{m=1}^{\infty} \frac{E[Y_m^2]}{m^2} < \infty.$

Proof. By Lemma 7.2,

$$E\left[Y_m^2\right] = \int_0^{\infty} 2yP(Y_m > y)dy = \int_0^m 2yP(Y_m > y)dy \leq \int_0^m 2yP(X_1 > y)dy,$$

so Tonelli's theorem gives

$$\sum_{m=1}^{\infty} \frac{E[Y_m^2]}{m^2} \leq \sum_{m=1}^{\infty} m^{-2} \int_0^m 2yP(X_1 > y)dy = 2\int_0^{\infty} \left(y \sum_{m>y} m^{-2} \right) P(X_1 > y)dy.$$

Since $\displaystyle\int_0^{\infty} P(X_1 > y)dy = E[X_1] < \infty$, we will be done if we can show that $\displaystyle y\sum_{m>y} m^{-2}$ is uniformly bounded.

To see that this is the case, observe that

$$y \sum_{m>y} m^{-2} \leq \sum_{m=1}^{\infty} m^{-2} = \frac{\pi^2}{6} < 2$$

for $y \in [0, 1]$, and for $j \geq 2$,

$$\sum_{m=j}^{\infty} m^{-2} \leq \int_{j-1}^{\infty} x^{-2}dx = (j-1)^{-1},$$

so

$$y \sum_{m>y} m^{-2} = y \sum_{m=\lfloor y \rfloor +1}^{\infty} m^{-2} \leq \frac{y}{\lfloor y \rfloor} \leq 2$$

for $y > 1$. □

It follows that $\displaystyle\sum_{n=1}^{\infty} P\left(\left|T_{k(n)} - E\left[T_{k(n)}\right]\right| > \varepsilon k(n)\right) < \infty$, so, since $\varepsilon > 0$ is arbitrary, the first Borel-Cantelli lemma implies that $\dfrac{T_{k(n)} - E\left[T_{k(n)}\right]}{k(n)} \to 0$ a.s.

Now $\displaystyle\lim_{k\to\infty} E[Y_k] = E[X_1]$ by the dominated convergence theorem, so $\lim_{n\to\infty} \dfrac{E\left[T_{k(n)}\right]}{k(n)} = E[X_1]$.

Thus we have shown that $\dfrac{T_{k(n)}}{k(n)} \to \mu$ almost surely.

Finally, if $k(n) \leq m < k(n+1)$, then

$$\frac{k(n)}{k(n+1)} \cdot \frac{T_{k(n)}}{k(n)} = \frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)} = \frac{T_{k(n+1)}}{k(n+1)} \cdot \frac{k(n+1)}{k(n)}$$

since T_n is nondecreasing.

Because $\dfrac{k(n+1)}{k(n)} = \dfrac{\lfloor \alpha^{n+1} \rfloor}{\lfloor \alpha^n \rfloor} \to \alpha$ as $n \to \infty$, we see that

$$\frac{\mu}{\alpha} \leq \liminf_{n\to\infty} \frac{T_m}{m} \leq \limsup_{n\to\infty} \frac{T_m}{m} \leq \alpha\mu,$$

and we're done since $\alpha > 1$ is arbitrary. □

The next result shows that the strong law holds whenever $E[X_1]$ exists.

Theorem 9.2. *Let $X_1, X_2, ...$ be i.i.d. with $E\left[X_1^+\right] = \infty$ and $E\left[X_1^-\right] < \infty$. Then $\frac{1}{n}S_n \to \infty$ a.s.*

Proof. For any $M \in \mathbb{N}$, let $X_i^M = X_i \wedge M$. Then the $X_i^M{}'s$ are i.i.d. with $E\left|X_1^M\right| < \infty$, so, writing $S_n^M = \sum_{i=1}^n X_i^M$, it follows from Theorem 9.1 that $\frac{1}{n}S_n^M \to E\left[X_1^M\right]$ almost surely as $n \to \infty$.
Now $X_i \geq X_i^M$ for all M, so $\liminf_{n\to\infty} \dfrac{S_n}{n} \geq \lim_{n\to\infty} \dfrac{S_n^M}{n} = E\left[X_1^M\right]$.
The monotone convergence theorem implies that

$$\lim_{M\to\infty} E\left[\left(X_1^M\right)^+\right] = E\left[\lim_{M\to\infty} \left(X_1^M\right)^+\right] = E\left[X_1^+\right] = \infty,$$

so

$$E\left[X_1^M\right] = E\left[\left(X_1^M\right)^+\right] - E\left[\left(X_1^M\right)^-\right] = E\left[\left(X_1^M\right)^+\right] - E\left[X_1^-\right] \nearrow \infty,$$

thus $\liminf_{n\to\infty} \frac{S_n}{n} \geq \infty$ a.s. and the theorem follows. □

Our first application of the strong law of large numbers comes from renewal theory.

Example 9.1. Let $X_1, X_2, ...$ be i.i.d. with $0 < X_1 < \infty$., and let $T_n = X_1 + ... + X_n$. Here we are thinking of the $X_i's$ as times between successive occurrences of events and T_n as the time until the nth event occurs. For example, consider a janitor who replaces a light bulb the instant it burns out. The first bulb is put in at time 0 and X_i is the lifetime of the ith bulb. Then T_n is the time that the nth bulb burns out and $N_t = \sup\{n : T_n \leq t\}$ is the number of light bulbs that have burned out by time t.

Theorem 9.3 (Elementary Renewal Theorem)**.** *If $E[X_1] = \mu \leq \infty$, then $\dfrac{N_t}{t} \to \dfrac{1}{\mu}$ a.s. as $t \to \infty$ (with the convention that $\frac{1}{\infty} = 0$).*

Proof. Theorems 9.1 and 9.2 imply that $\lim_{n\to\infty} \dfrac{T_n}{n} = \mu$ a.s., and it follows from the definition of N_t that $T_{N_t} \leq t < T_{N_t+1}$, hence

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t}.$$

Since $T_n < \infty$ for all n, we have that $N_t \nearrow \infty$ as $t \nearrow \infty$. Thus there is a set Ω_0 with $P(\Omega_0) = 1$ such that $\lim_{n\to\infty} \dfrac{T_n(\omega)}{n} = \mu$ and $\lim_{t\to\infty} N_t(\omega) = \infty$, hence

$$\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)} \to \mu, \quad \frac{N_t(\omega)+1}{N_t(\omega)} \to 1,$$

for all $\omega \in \Omega_0$.
It follows that $\dfrac{t}{N_t} \to \mu$ on Ω_0, which implies the result. □

Example 9.2. A common situation in statistics is that one has a sequence of random variables which is assumed to be i.i.d., but the underlying distribution is unknown. A popular estimate for the true distribution function $F(x) = P(X_1 \leq x)$ is given by the empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_i).$$

That is, one approximates the true probability of being at most x with the observed frequency of values $\leq x$ in the sample. The strong law provides some justification for this method of inference by showing that for every $x \in \mathbb{R}$, $F_n(x) \to F(x)$ almost surely as $n \to \infty$. The next result shows that the convergence is actually uniform in x.

Theorem 9.4 (Glivenko-Cantelli). *As $n \to \infty$*

$$\sup_x |F_n(x) - F(x)| \to 0 \ a.s.$$

Proof.

Fix $x \in \mathbb{R}$ and let $Y_n = 1\{X_n < x\}$. Then $Y_1, Y_2, ...$ are i.i.d. with $E[Y_1] = P(X_1 < x) = F(x^-)$, so the strong law implies that $F_n(x^-) = \frac{1}{n}\sum_{i=1}^n Y_i \to F(x^-)$ a.s. as $n \to \infty$. Similarly, $F_n(x) \to F(x)$ a.s.

In general, for any countable collection $\{x_i\} \subseteq \mathbb{R}$, there is a set Ω_0 with $P(\Omega_0) = 1$ such that $F_n(x_i)(\omega) \to F(x_i)$ and $F_n(x_i^-)(\omega) \to F(x_i^-)$ for all $\omega \in \Omega_0$.

For each $k \in \mathbb{N}$, $j = 1, ..., k-1$, set $x_{j,k} = \inf\left\{y : F(y) \geq \frac{j}{k}\right\}$. The pointwise convergence of $F_n(x)$ and $F_n(x^-)$ implies that we can pick $N_k(\omega) \in \mathbb{N}$ such that

$$\left|F_n({x_{j,k}}^-)(\omega) - F({x_{j,k}}^-)\right|, |F_n(x_{j,k})(\omega) - F(x_{j,k})| < \frac{1}{k} \text{ for all } j = 1, ..., k-1$$

whenever $n \geq N_k(\omega)$. Setting $x_{0,k} := -\infty$ and $x_{k,k} := +\infty$, we see that the above inequalities also hold for $j = 0, k$.

Thus if $x_{j-1,k} < x < x_{j,k}$ with $1 \leq j \leq k$ and $n \geq N_k$, then the inequality $F({x_{j,k}}^-) - F(x_{j-1,k}) < \frac{1}{k}$ and the monotonicity of F_n and F imply

$$F_n(x) \leq F_n({x_{j,k}}^-) \leq F({x_{j,k}}^-) + \frac{1}{k} \leq F(x_{j-1,k}) + \frac{2}{k} \leq F(x) + \frac{2}{k},$$
$$F_n(x) \geq F_n(x_{j-1,k}) \geq F(x_{j-1,k}) - \frac{1}{k} \geq F({x_{j,k}}^-) - \frac{2}{k} \geq F(x) - \frac{2}{k}.$$

Consequently, we have $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leq \frac{2}{k}$ and the theorem follows. □