STATS 310A: Theory of Probability I	**Autumn 2016/17**

Lecture 13: November 7

Lecturer: Persi Diaconis — *Scribes: Kenneth Tay*

13.1 Strong Law of Large Numbers

Theorem 13.1 (Kolmogorov) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_i, i = 1, 2, \dots$ be i.i.d. random variables with finite mean $\mathbb{E}(X_1) = \mu < \infty$.*

If $S_n = X_1 + \cdots + X_n$, then $\dfrac{S_n}{n} \to \mu$ almost surely.

Proof:[Etemadi] We break the proof into a number of steps.

1. **Confine to non-negative random variables.** Since $X_1 = X_1^+ - X_1^-$, $\mu = \mu^+ - \mu^-$, it's enough to prove the theorem for non-negative random variables.

2. **Truncation.** Let $Y_i = X_i \delta_{\{X_i \leq i\}}$. We will prove the Law of Large Numbers for $\{Y_i\}$ first, then try to go back to $\{X_i\}$.

3. **Take subsequences.** Let $\alpha > 1$, $u_n = \lfloor \alpha^n \rfloor$. We claim that for any $\varepsilon > 0$,

$$\sum_{n=1}^\infty P\left\{ \left| \frac{T_{u_n} - \mathbb{E}T_{u_n}}{u_n} \right| \geq \varepsilon \right\} < \infty \tag{13.1}$$

where $T_k = Y_1 + \cdots + Y_k$.

Note that

$$\begin{aligned} \text{Var } T_n &= \sum_{k=1}^n \text{Var } (Y_k) \\ &\leq \sum_{k=1}^n \mathbb{E}\left[X_1^2 \delta_{\{X_1 \leq k\}} \right] \\ &\leq n\mathbb{E}\left[X_1^2 \delta_{\{X_1 \leq n\}} \right]. \end{aligned}$$

Thus, by Chebyshev's inequality,

$$\begin{aligned} \sum_{n=1}^\infty P\left\{ \left| \frac{T_{u_n} - \mathbb{E}T_{u_n}}{u_n} \right| \geq \varepsilon \right\} &\leq \sum_{n=1}^\infty \frac{\text{Var } T_{u_n}}{\varepsilon^2 u_n^2} \\ &\leq \sum_{n=1}^\infty \frac{1}{\varepsilon^2 u_n^2} \cdot u_n \mathbb{E}\left[X_1^2 \delta_{\{X_1 \leq u_n\}} \right] \\ &= \frac{1}{\varepsilon^2} \mathbb{E}\left[X_1^2 \sum_{n=1}^\infty \frac{1}{u_n} \delta_{\{X_1 \leq u_n\}} \right]. \end{aligned}$$

Set $k = \dfrac{2\alpha}{\alpha - 1}$ and fix any $x > 0$. Let n_x be the smallest integer such that $u_{n_x} \geq x$. It is easy to check that $\alpha^{n_x} \geq x$. We thus have

$$\sum_{u_n \geq x} \frac{1}{u_n} \leq 2 \sum_{n \geq n_x} \alpha^{-n} = k\alpha^{-n_x} \leq \frac{k}{x},$$

so

$$\sum_{n=1}^{\infty} \frac{1}{u_n} \delta_{\{X_1 \leq u_n\}} \leq \frac{k}{X_1}$$

for $X_1 > 0$. Thus,

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{T_{u_n} - \mathbb{E}T_{u_n}}{u_n}\right| \geq \varepsilon\right\} \leq \frac{k}{\varepsilon^2}\mathbb{E}\left[X_1\right] < \infty.$$

(Note that the conclusion still holds even if $X_1 = 0$ at any point.) For any ε, Equation 13.1 is true. By Borel-Cantelli, $\dfrac{T_{u_n} - \mathbb{E}T_{u_n}}{u_n} > \varepsilon$ happens only finitely often. Taking $\varepsilon = \dfrac{1}{k}$ for $k \in \mathbb{N}$ and letting $k \to \infty$, we have $\dfrac{T_{u_n} - \mathbb{E}T_{u_n}}{u_n} \to 0$ almost surely.

Consider the following "baby fact of analysis":

Lemma 13.2 *If $x_n \to x$, then $\frac{1}{n}\sum_{i=1}^{n} x_i \to x$.*

We know that $\mathbb{E}Y_i = \mathbb{E}X_1\delta_{\{X_1 \leq i\}} \nearrow \mathbb{E}X_1$, so by the lemma, $\mathbb{E}[T_n/n] \to \mu$. Thus we now have $T_{u_n}/u_n \to \mu$ almost surely.

(At this point, we have the theorem we want for truncated variables on subsequences.)

4. **Remove truncation.** Consider $\sum_{n=1}^{\infty} P\{X_n \neq Y_n\}$:

$$\sum_{n=1}^{\infty} P\{X_n \neq Y_n\} = \sum_{n=1}^{\infty} P\{X_1 \geq n\} \leq \int_0^{\infty} P\{X_1 > t\}dt = \mu,$$

which is finite. Therefore by Borel-Cantelli, $X_n \neq Y_n$ only happens finitely often, so

$$\frac{S_n - T_n}{n} \xrightarrow{a.s.} 0, \quad \Rightarrow \quad \frac{S_{u_n}}{u_n} \xrightarrow{a.s.} \mu.$$

5. **Interpolate (i.e. remove subsequences).** For any k, find n such that $u_n \leq k < u_{n+1}$. Then

$$\frac{S_{u_n}}{u_n}\frac{u_n}{u_{n+1}} \leq \frac{S_k}{k} \leq \frac{S_{u_{n+1}}}{u_{n+1}}\frac{u_{n+1}}{u_n}.$$

Note that $u_{n+1}/u_n \to \alpha$, so letting $n \to \infty$, we have

$$\frac{1}{\alpha}\mu \leq \liminf \frac{S_k}{k} \leq \limsup \frac{S_k}{k} \leq \alpha\mu$$

almost surely. Since this holds for all $\alpha > 1$, by letting α go to 1, we must have $\lim S_n/n = \mu$.

■

Comments:

1. We only used independence for

   $$\text{the variance of sum} = \text{sum of the variances.}$$

   Thus, if X_i are identically distributed and pairwise independent, we still have $S_n/n \to \mu$.

2. This is a "4 T's proof": Truncation, Tchebychev, inTerpolation, and Tubsequences.

3. The Strong Law of Large Numbers is a special case of the Martingale Convergence Theorem and the Ergodic Theorem.

4. (Complaint) It is an amazing, clean statement: $S_n/n \to \mu$ a.s.

   But what's the real content of this statement? It says that S_n/n gets close to μ and stays there.

   What we would like instead is some sort of quantitative bound, e.g.

   $$P\left\{\left|\frac{S_n}{n} - \mu\right| < \varepsilon \text{ for all } n \geq N\right\} \geq 1 - f(N, \varepsilon),$$

   (the first time that S_n/n is close to μ and stays there).

The corollary below is a converse of sorts:

Corollary 13.3 *Let X_i, $1 \leq i < \infty$, be i.i.d. random variables with $\mathbb{E}X^- < \infty$, $\mathbb{E}X^+ = \infty$. Then $\mathbb{E}X = \infty$, and $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \infty$.*

Proof: By the Strong Law of Large Numbers on X^-,

$$\frac{1}{n}\sum_{1}^{n} X_i^- \xrightarrow{a.s.} \mathbb{E}(X_1^-) < \infty.$$

Let $X_k^R = X_n^+ \delta_{\{X_n \leq R\}}$. Then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \geq \frac{1}{n}\sum X_k^R \xrightarrow{a.s.} \mathbb{E}(X_k^R).$$

Letting $R \to \infty$, we have the desired result. ∎

Theorem 13.4 (Siegmund) *Let X_i, $1 \leq i < \infty$, be i.i.d., with mean 0 and variance 1. Let*

$$m(\varepsilon) = m = \sup\left\{n \geq 0 : \left|\frac{S_n}{n}\right| \geq \varepsilon\right\}.$$

Then, for $0 \leq x < \infty$,

$$P\{\varepsilon^2 m \leq x\} \to 2\Phi(x) - 1$$

as $\varepsilon \to 0$, where $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ is the normal distribution function.

(In other words, the parameter m scales like $\frac{1}{\varepsilon^2}$.)

Example (Cauchy): If X_1 has density $\dfrac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$, then $P\{S_n/n \leq x\} = P(X_1 \leq x)$ for all n.