Math 5652: Introduction to Stochastic Processes: Spring 2014

Appendix C. Strong Law of Large Numbers.

The strong law of large numbers is formulated (without proof) in Sec. 3.1, Theorem 3.2 of the textbook [D]. In this note, we give a complete proof of this fact. For further information, see [F], Ch.7.

Proposition 1 (Markov's Inequality). *Let $Y \geq 0$ be a random variable. Then*

$$P(Y \geq a) \leq \frac{E(Y)}{a} \quad \text{for any constant} \quad a > 0. \tag{1}$$

Proof. For fixed $a = \text{const} > 0$, consider the event $A := \{\omega \in \Omega : \ X(\omega) \geq a > 0\}$ and its indicator

$$I_A(\omega) := \begin{cases} 1 & \text{if} \quad \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Since $X \geq a\, I_A$, we have

$$E(X) \geq a \cdot E(I_A) = a \cdot P(A) = a \cdot P(X \geq a),$$

and (1) follows. □

Proposition 2 (Chebyshev's's Inequality). *Let X be a random variable with $\mu = E(X)$ and $\mathrm{Var}\,(X) < \infty$. Then*

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}\,(X)}{\varepsilon^2} \quad \text{for any constant} \quad \varepsilon > 0. \tag{3}$$

Proof. Set $Y := |X - \mu|^2$, $a := \varepsilon^2 > 0$. Then by Markov's inequality,

$$P(|X - \mu| \geq \varepsilon) = P(Y \geq a) \leq \frac{E(Y)}{a} = \frac{\mathrm{Var}\,(X)}{\varepsilon^2}.$$

□

Theorem 3 (Weak Law of Large Numbers). *Let $X_1, X_2, \dots, X_n, \dots$ be independent identically distributed (i.i.d.) random variables with $\mu = E(X)$ and $\mathrm{Var}\,(X) < \infty$. Then the* **sample mean**

$$\overline{X_n} := \frac{1}{n}\,(X_1 + X_2 + \cdots + X_n) \to \mu \quad \textbf{in probability} \text{ as } \quad n \to \infty,$$

i.e. for any constant $\varepsilon > 0$,

$$P(|\overline{X_n} - \mu| \geq \varepsilon) \to 0 \quad \text{as} \quad n \to \infty. \tag{4}$$

Proof. We have

$$E(\overline{X_n}) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu, \qquad \mathrm{Var}\,(\overline{X_n}) = \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}\,(X_k) = \frac{\sigma^2}{n}.$$

Therefore, by Chebyshev's Inequality,

$$P(|\overline{X_n} - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}\,(\overline{X_n})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as} \quad n \to \infty.$$

□

Lemma 4 (Borel-Cantelli's Lemma). *Let $A_1, A_2, \dots$ be a sequence of events such that $\sum P(A_n) < \infty$. Then with probability one only finitely many events A_n occur. In other words, the event*

$$A := \limsup_{n\to\infty} A_n := \left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \right) \quad \text{has probability} \quad P(A) = 0. \tag{5}$$

Proof. For every natural k,

$$0 \le P(A) \le P\Big(\bigcup_{n=k}^{\infty} A_n \Big) \le \sum_{n=k}^{\infty} P(A_n).$$

Since $P(A)$ does not depend on k, and the right hand side converges to 0 as $k \to \infty$, we must have $P(A) = 0$. □

The Strong Law of Large Numbers (Theorem 7 below) together with two preparatory Theorems 5 and 6, are due to A.N. Kolmogorov.

Theorem 5 (Kolmogorov's Inequality). *Let $X_1, X_2, \dots, X_n$ be independent random variables with $E(X_k) = 0$ and $\operatorname{Var}(X_k) = \sigma_k^2$ for all $k = 1, 2, \dots, n$. Then $S_k := X_1 + X_2 + \cdots + X_k$ for $k = 1, 2, \dots, n$ satisfy*

$$P\Big(\max_{1\le k\le n} |S_k| \ge t a_n \Big) \le \frac{1}{t^2} \quad \text{for any constant} \quad t > 0, \tag{6}$$

where $a_n^2 := \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 = \operatorname{Var}(S_n) = E(S_n^2)$.

Proof. Introduce the stopping time

$$T := \min\{k \ge 1 : \ |S_k| \ge t a_n\} \quad \text{if} \quad \max_{1\le k\le n} |S_k| \ge t a_n,$$

and $T = n$ otherwise. By Markov's inequality (1) with $Y := S_T^2$ and $a := t^2 a_n^2$,

$$P(|S_T| \ge t a_n) = P(|S_T^2| \ge t^2 a_n^2) \le \frac{E(S_T^2)}{t^2 a_n^2} \tag{7}$$

We can write

$$E(S_T^2) = \sum_{k=1}^{n} E(I_k S_k^2) \quad \text{where} \quad I_k := I_{\{T=k\}}. \tag{8}$$

Further

$$S_n^2 = \big[S_k + (S_n - S_k) \big]^2 = S_k^2 + 2 S_k (S_n - S_k) + (S_n - S_k)^2. \tag{9}$$

Since $I_k S_k$ and $S_n - S_k = X_{k+1} + \cdots + X_n$ are independent and $E(S_n - S_k) = 0$, we obtain

$$E\big(I_k S_k (S_n - S_k) \big) = E(I_k S_k) \cdot E(S_n - S_k) = 0 \quad \text{for all} \quad k.$$

Then from (9) it follows $E(I_k S_n^2) \ge E(I_k S_k^2)$ for $k = 1, 2, \dots, n$. Together with (8), these imply

$$E(S_T^2) \le \sum_{k=1}^{n} E(I_k S_n^2) = E(S_n^2) = a_n^2,$$

and the desired inequality (6) follows from (7) by definition of T. □

Theorem 6 (Kolmogorov's Test)**.** *Let $X_1, X_2, \dots, X_n, \dots$ be independent random variables with $E(X_k) = 0$ and $\operatorname{Var}(X_k) = \sigma_k^2$ for all $k = 1, 2, \dots, n, \dots,$ such that $\sum k^{-2}\sigma_k^2 < \infty$. Then*

$$\frac{S_n}{n} := \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \quad \text{almost surely (a.s.)}, \quad \text{i.e.} \quad P\biggl(\frac{S_n}{n} \to 0 \quad \text{as} \quad n \to \infty\biggr) = 1. \tag{10}$$

Proof. Fix an arbitrary $\varepsilon > 0$. For $m = 0, 1, 2, \dots,$ consider the events

$$A_m := \biggl\{ \max_{2^{m-1} < k \le 2^m} \frac{|S_k|}{k} \ge \varepsilon \biggr\} \subset B_m := \biggl\{ \max_{1 \le k \le 2^m} |S_k| \ge 2^{m-1}\varepsilon \biggr\}.$$

We can apply Theorem 5 with $n = 2^m$ and $t = 2^{m-1}\varepsilon a_n^{-1}$. This gives us

$$P(A_m) \le P(B_m) \le \frac{1}{t^2} = \frac{4a_n^2}{\varepsilon^2 n^2},$$

which in turn implies

$$\sum_{m=0}^{\infty} P(A_m) \le \frac{4}{\varepsilon^2} \sum_{m=0}^{\infty} \frac{1}{4^m} \sum_{k=1}^{2^m} \sigma_k^2.$$

The right hand side can be considered as the double sum over all the integers $m \ge 0$ and $k \ge 1$ satisfying $1 \le k \le 2^m$. Changing the order of summation, we write

$$\sum_{m=0}^{\infty} P(A_m) \le \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{m=m_0}^{\infty} \frac{1}{4^m},$$

where $m_0 = m_0(k)$ is the minimal integer m satisfying $k \le 2^m$, so that $k \le 2^{m_0} < 2k$. The sum of the geometric series

$$\sum_{m=m_0}^{\infty} \frac{1}{4^m} = \frac{4}{3 \cdot 4^{m_0}} \le \frac{4}{3k^2}.$$

Hence

$$\sum_{m=0}^{\infty} P(A_m) \le \frac{16}{3\varepsilon^2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty.$$

By Lemma 4, with probability one only finite number of A_m occurs, which means that

$$P\biggl(\limsup_{n\to\infty} \frac{|S_n|}{n} \le \varepsilon \biggr\} = 1.$$

Since $\varepsilon > 0$ is arbitrary, the desired property (10) follows. □

Theorem 7 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \dots, X_n, \dots$ be i.i.d. random variables with $E(|X|) < \infty$ and $\mu = E(X)$. Then*

$$\overline{X_n} := \frac{1}{n}(X_1 + \cdots + X_n) \to \mu \quad \text{a.s.} \quad \text{as} \quad n \to \infty. \tag{11}$$

Proof. Replacing X_k by $X_k - \mu$ and $\overline{X_n}$ by $\overline{X_n} - \mu$, we reduce the proof to the case $\mu = 0$. For $k = 1, 2, \dots,$ represent X_k in the form

$$X_k = U_k + V_k, \quad \text{where} \quad U_k := I_{\{|X_k| < k\}} \cdot X_k, \quad V_k := I_{\{|X_k| \ge k\}} \cdot X_k. \tag{12}$$

Denote $\mu_k := E(U_k)$. Since $E(X_k) = 0$, we have $E(V_k) = E(X_k - U_k) = -\mu_k$, and

$$\left|\sum_{k=1}^{n} \mu_k\right| = \left|\sum_{k=1}^{n} E(V_k)\right| \le \sum_{k=1}^{n} E\big(|V_k|\big) = \sum_{k=1}^{n} E\Big(I_{\{|X_k|\ge k\}} \cdot |X_k|\Big).$$

Here distributions of X_k do not depend on k, so that the last expression can be rewritten as

$$\sum_{k=1}^{n} E\Big(I_{\{|X|\ge k\}} \cdot |X|\Big) = E\left(\sum_{k=1}^{n} I_{\{|X|\ge k\}} \cdot |X|\right) \le E\big(\min\{n, |X|\} \cdot |X|\big).$$

By the Monotone Convergence Theorem in Real Analysis,

$$\left|\frac{1}{n}\sum_{k=1}^{n} \mu_k\right| \le \frac{1}{n} \cdot E\big(\min\{n, |X|\} \cdot |X|\big) = E\left(\min\left\{1, \frac{|X|}{n}\right\} \cdot |X|\right) \to 0 \quad \text{as} \quad n \to \infty. \tag{13}$$

Further, introduce the quantities

$$c_j := E\Big(I_{\{j-1\le|X|<j\}} \cdot |X|\Big) \quad \text{for} \quad j = 1, 2, \ldots.$$

Then

$$\sigma_k^2 := \mathrm{Var}\,(U_k) \le E(U_k^2) = E\Big(I_{\{|X|<k\}} \cdot |X|^2\Big) = \sum_{j=1}^{k} E\Big(I_{\{j-1\le|X|<j\}} \cdot |X|^2\Big) \le \sum_{j=1}^{k} j\, c_j.$$

Note that

$$\sum_{k=j}^{\infty} \frac{1}{k^2} \le \sum_{k=j}^{\infty} \frac{2}{k(k+1)} = 2 \cdot \sum_{k=j}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \frac{2}{j} \quad \text{for} \quad j = 1, 2, \ldots.$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \le \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{j=1}^{k} j\, c_j = \sum_{1\le j\le k} \frac{j\, c_j}{k^2} = \sum_{j=1}^{\infty} j\, c_j \sum_{k=j}^{\infty} \frac{1}{k^2} \le 2 \cdot \sum_{j=1}^{\infty} c_j = 2 \cdot E(|X|) < \infty.$$

By Theorem 6 applied to $U_k - \mu_k$ instead of X_k, we get

$$\frac{1}{n}\sum_{k=1}^{n} (U_k - \mu_k) \to 0 \quad \text{a.s} \quad \text{as} \quad n \to \infty. \tag{14}$$

In addition, we have

$$\sum_{k=1}^{\infty} P\big(|X_k| \ge k\big) = \sum_{k=1}^{\infty} E\Big(I_{\{|X|\ge k\}}\Big) = E\left(\sum_{k=1}^{\infty} I_{\{|X|\ge k\}}\right) \le E(|X|) < \infty.$$

By the Borel-Cantelli Lemma (Lemma 4), only finitely many events $\{|X_k| \ge k\}$ occur (a.s.). This means that in (12) $X_k = U_k$ except for finitely many indices k. In combination with (13) and (14), this implies the desired property (11):

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} U_k = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} (U_k - \mu_k) = 0 \quad \text{(a.s.)} \quad .$$

Theorem is proved. □

References

[D] Richard Durrett, *Essentials of Stochastic Processes*, 2nd Edition, Springer, 2012.

[F] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, John Wiley & Sons, Inc., 1971.