

STAT 7200

Introduction to Advanced Probability

Lecture 12

Taylor R. Brown

- Almost Sure Convergence
- Convergence in Probability
- Law of Large Numbers

Almost Sure Convergence

- We say that $\{Z_n\}$ converges to Z almost surely (or a.s., or with probability 1), if $P(\{\omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega), \omega \in \Omega\}) = 1$
- $P(Z_n \rightarrow Z) = 1$ is equivalent to
For each $\varepsilon > 0$, $P(|Z_n - Z| \geq \varepsilon \text{ i.o.}) = 0$ (or
 $P(|Z_n - Z| < \varepsilon \text{ a.a.}) = 1$)
- For r.v.s. Z, Z_1, Z_2, \dots , we have that $\varepsilon > 0$, $\sum_n P(|Z_n - Z| \geq \varepsilon) < \infty$ implies $P(Z_n \rightarrow Z) = 1$ by the Borel-Cantelli lemma.

Convergence in Probability

- **Convergence in Probability** For r.v.s. Z, Z_1, Z_2, \dots , we say that $\{Z_n\}$ converges to Z in probability, if for all $\varepsilon > 0$, $P(|Z_n - Z| \geq \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$.
- **Example** Let Z_1, Z_2, \dots be random variables such that $P(Z_n = 1) = \frac{1}{2^n}$ and $P(Z_n = 0) = 1 - \frac{1}{2^n}$. Then for $1 > \varepsilon > 0$, $P(|Z_n| \geq \varepsilon) = \frac{1}{2^n} \rightarrow 0$, so we must have $Z_n \rightarrow 0$ in probability.

Proposition 1

Almost Sure Convergence implies convergence in probability.

- **Proof** If $\{Z_n\}$ converges to Z almost surely, then for each $\varepsilon > 0$, $P(|Z_n - Z| \geq \varepsilon \text{ i.o.}) = 0$. That is, $P(\limsup_n \{|Z_n - Z| \geq \varepsilon\}) = 0$.
- $0 \leq \liminf_n P(|Z_n - Z| \geq \varepsilon) \leq \limsup_n P(|Z_n - Z| \geq \varepsilon) \leq P(\limsup_n \{|Z_n - Z| \geq \varepsilon\}) = 0$.
- Thus, $\lim_n P(|Z_n - Z| \geq \varepsilon) = 0$, $\{Z_n\}$ converges to Z in probability.

Convergence in Probability and Almost Sure Convergence

On the other hand, even if we know $P(|Z_n - Z| \geq \varepsilon) \rightarrow 0$, we still have no idea regarding the probability of $\{|Z_n - Z| \geq \varepsilon \text{ i.o.}\}$ is greater than 0.

If $P(|Z_n - Z| \geq \varepsilon)$ went to 0 fast enough, then the $\sum_n P(|Z_n - Z| \geq \varepsilon) < \infty$, and we could invoke Borel-Cantelli. If it goes to 0 slowly, then we can come up with counter-examples:

Convergence in Probability and Almost Sure Convergence

- For $n \in \mathbb{N}$ let Z_n be such that $P(Z_n = 1) = 1 - P(Z_n = 0) = \frac{1}{n}$. Also, suppose these r.v.s are independent.
- Pick $1 > \epsilon > 0$. Then $P(|Z_n - 0| > \epsilon) = P(Z_n = 1) = n^{-1} \rightarrow 0$. So it converges *in probability*.
- On the other hand, $\sum_n P(|Z_n - 0| > \epsilon) = \infty$, so by the second Borel-Cantelli lemma, $P(|Z_n| > \epsilon \text{ i.o.}) = 1$. Not only does it not converge almost surely to 0, but it converges to 0 almost nowhere!

Convergence in Probability and Almost Sure Convergence

- **Example** Consider the uniform measure $([0, 1], \mathcal{M}, \lambda)$ and define

$$Z_1 = 1_{[0, 1/2)}, Z_2 = 1_{[1/2, 1)},$$

$$Z_3 = 1_{[0, 1/4)}, Z_4 = 1_{[1/4, 1/2)}, Z_5 = 1_{[1/2, 3/4)}, Z_6 = 1_{[3/4, 1)},$$

$$Z_7 = 1_{[0, 1/8)}, Z_8 = 1_{[1/8, 2/8)}, \dots, Z_{14} = 1_{[7/8, 1)},$$

...

- In general $Z_n = 1_{[\frac{k}{2^m}, \frac{k+1}{2^m})}$ where $m = \lfloor \log_2(n+1) \rfloor$ and $k = n + 1 - 2^m$
- $\{Z_n\}$ converges to 0 in probability because, for any $1 > \epsilon > 0$,
 $2^{-\lfloor \log_2(n+1) \rfloor} \leq 1/(n+1) \rightarrow 0$.
- But $\limsup_{n \rightarrow \infty} \{|Z_n| > \epsilon\} = \bigcap_{a=1}^{\infty} \bigcup_{n \geq a} [\frac{k}{2^m}, \frac{k+1}{2^m}) = \bigcap_{a=1}^{\infty} [0, 1]$.
- For each ω , the sequence $Z_1(\omega), Z_2(\omega), \dots$ contains infinitely number of 1s.

Weak Law of Large Numbers Version 1

Theorem 2 (WLLN V1)

For a sequence of independent random variables X_1, X_2, \dots with the same mean μ and finite variance bounded by σ^2 , define $S_n = X_1 + X_2 + \dots + X_n$, then S_n/n converges to μ in probability.

- **Proof** We need to prove that, for any $\varepsilon > 0$,
 $\lim_n P(|S_n/n - \mu| \geq \varepsilon) = 0$.
- Since $E(S_n/n) = \sum_{i=1}^n E(X_i)/n = \mu$, by Chebychev's inequality, we have:

$$P(|\frac{S_n}{n} - \mu| \geq \varepsilon) \leq \frac{\text{Var}(S_n/n)}{\varepsilon^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2 \varepsilon^2} = \frac{\sigma^2}{n \varepsilon^2} \rightarrow 0$$

- **Note** Since $\sum_n \frac{1}{n} = \infty$, the above proof does not suggest that S_n/n converges to μ almost surely.

Strong Law of Large Numbers Version 1

Theorem 3 (SLLN V1)

For a sequence of independent random variables X_1, X_2, \dots with the same mean μ and bounded finite fourth central moments ($E(X_i - \mu)^4 \leq a < \infty$), define $S_n = X_1 + X_2 + \dots + X_n$, then S_n/n converges to μ almost surely.

- **Proof** Without loss of generality, let us assume that $\mu = 0$ (otherwise we can set $X'_i = X_i - \mu$).
- We want to show that, for any $\varepsilon > 0$, $\sum_n P(|S_n/n| \geq \varepsilon) < \infty$, and then invoke BCL. Note that

$$P(|\frac{S_n}{n}| \geq \varepsilon) = P(S_n^4 \geq n^4 \varepsilon^4) \leq \frac{E(S_n^4)}{n^4 \varepsilon^4}$$

If we can show that, $E(S_n^4) \leq Kn^2$, where K is a constant, then $\sum_n P(|S_n/n| \geq \varepsilon) \leq K\varepsilon^{-4} \sum_n \frac{1}{n^2} < \infty$, and we should have S_n/n converges to μ almost surely.

Strong Law of Large Number Version 1: continued

- **Proof: continued** As $S_n = X_1 + X_2 + \cdots + X_n$, the expansion of S_n^4 would contains four different terms: 1) X_i^4 , whose expectation is bounded by constant a ; 2) $X_i(X_j^3)$, whose expectation equals 0 as we assume $\mu = 0$; 3) $X_i X_j X_k^2$, whose expectation also equals 0. 4) $X_i^2 X_j^2$.
 - For the expectation of $X_i^2 X_j^2$ ($i \neq j$). Note that as $X^2 \leq X^4 + 1$ (considering $X > 1$ and $X \leq 1$), we have $EX_i^2 \leq EX_i^4 + 1 \leq a + 1$, so $E(X_i^2 X_j^2) \leq (a + 1)^2$.
 - Furthermore, there are n different terms in the form of X_j^4 in the expansion of S_n , and $\binom{4}{2} \binom{n}{2} = 3n(n - 1)$ different terms in the form of $X_i^2 X_j^2$. Thus:

$$E(S_n^4) = \sum_i E(X_i^4) + \sum_{i \neq j} E(X_i^2 X_j^2) \leq na + 3n(n - 1)(a + 1)^2 \leq Kn^2.$$

We then have S_n/n converges to μ almost surely.

Relax the Conditions in the Law of Large Number

- In the discussion above, to ensure the law of large numbers, we required that X_i s are independent random variables with finite high-order moments.
- One way to relax this condition is to only require the first moment to be finite. In this situation, we would need to add an extra condition: that the X_i s are identically distributed (in addition to independent).
- **Identically Distributed:** A collection of random variable $\{X_\alpha\}_{\alpha \in I}$ is identically distributed if for any measurable function f , the expectation $E(f(X_\alpha))$ is the same for all $\alpha \in I$. This condition is equivalent to: for any $x \in \mathbb{R}$, $P(X_\alpha \leq x)$ does not depend on α .
- **i.i.d.:** A collection of random variable $\{X_\alpha\}_{\alpha \in I}$ are i.i.d. if they are independent and identically distributed.

Strong Law of Large Number Version 2

Theorem 4 (SLLN V2)

Let X_1, X_2, \dots be a sequence of i.i.d. random variables with a finite mean μ . Define $S_n = X_1 + X_2 + \dots + X_n$. Then S_n/n converges to μ almost surely.

Corollary 5 (Weak Law of Large Number (WLLN) V2)

Let X_1, X_2, \dots be a sequence of i.i.d. random variables with a finite mean μ . Define $S_n = X_1 + X_2 + \dots + X_n$. Then S_n/n converges to μ in probability.

The second version of WLLN follows from the fact that Almost Sure Convergence implies convergence in probability.

Proof to Strong Law of Large Number Version 2: Part I

- First, without loss of generality, we may assume that $X > 0$. Otherwise, we can let $X_i = X_i^+ - X_i^-$, and apply the law of large number to X_i^+ and X_i^- respectively.
- Second, to prove almost sure convergence, the most reliable route is to use Chebchev's inequality to obtain an upper bound of $P(|S_n/n - \mu| \geq \varepsilon)$ and then apply Borel-Cantelli Lemma to show that the probability of event $\{|S_n/n - \mu| \geq \varepsilon \text{ i.o.}\}$ equals 0. However, the condition of applying Chebchev's inequality is that the variance of X_i exists. For this purpose, we need to construct a truncated version of X_i .

Proof to Strong Law of Large Number Version 2: Part II

- Let $Y_i = X_i 1_{X_i \leq i}$. Then $0 \leq Y_i \leq i$, $E(Y_i^k) \leq i^k < \infty$ for any k .

Lemma 6

Define $T_n = Y_1 + \cdots + Y_n$, if T_n/n converges to μ almost surely, S_n/n also converges to μ almost surely

- **Proof:** We only need to show that $(T_n - S_n)/n \rightarrow 0$ almost surely.
 - As $\sum_{k=1}^{\infty} P(X_k \neq Y_k) = \sum_{k=1}^{\infty} P(X_k > k) \leq \sum_{k=1}^{\infty} P(X_1 \geq k) \leq E(X_1) = \mu < \infty$ (see Proposition 4.2.9), by the Borel-Cantelli Lemma, $P(X_k \neq Y_k \text{ i.o.}) = 0$. Thus $P(X_k - Y_k = 0 \text{ a.a.}) = 1$.
 - For any $\omega \in \{\omega : X_k(\omega) - Y_k(\omega) = 0 \text{ a.a.}\}$, there is an $N \in \mathbb{N}$ such that for any $n > N$, $X_n(\omega) = Y_n(\omega)$. Correspondingly, for $n > N$, $(T_n(\omega) - S_n(\omega))/n = \sum_{i=1}^N (Y_i(\omega) - X_i(\omega))/n \rightarrow 0$ as $n \rightarrow \infty$. Thus $P(\lim_n (T_n - S_n)/n = 0) \leq P(X_k - Y_k = 0 \text{ a.a.}) = 1$.