18.175: Lecture 6

Borel-Cantelli and strong law

Scott Sheffield

MIT

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1 + X_2 + \ldots + X_n}{n}$ is called the *empirical average* of the first n trials.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.
- Indeed, **weak law of large numbers** states that for all $\epsilon > 0$ we have $\lim_{n\to\infty} P\{|A_n - \mu| > \epsilon\} = 0$.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.
- Indeed, **weak law of large numbers** states that for all $\epsilon > 0$ we have $\lim_{n\to\infty} P\{|A_n - \mu| > \epsilon\} = 0$.
- Example: as n tends to infinity, the probability of seeing more than $.50001n$ heads in n fair coin tosses tends to zero.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.
- But weak law holds as long as $E[|X|]$ is finite, so that μ is well defined.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.
- But weak law holds as long as $E[|X|]$ is finite, so that μ is well defined.
- One standard proof uses characteristic functions.

Characteristic functions

- Let X be a random variable.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- But characteristic functions have an advantage: they are well defined at all t for all random variables X.

Continuity theorems

- Let X be random variable, X_n a sequence of random variables.

Continuity theorems

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.

Continuity theorems

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).

Continuity theorems

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).
- **Lévy's continuity theorem (coming later):** if
$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$$
for all t, then X_n converge in law to X.

Continuity theorems

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).
- **Lévy's continuity theorem (coming later):** if

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$$

  for all t, then X_n converge in law to X.
- By this theorem, we can prove weak law of large numbers by showing $\lim_{n\to\infty} \phi_{A_n}(t) = \phi_\mu(t) = e^{it\mu}$ for all t. When $\mu = 0$, amounts to showing $\lim_{n\to\infty} \phi_{A_n}(t) = 1$ for all t.
- **Moment generating analog:** if moment generating functions $M_{X_n}(t)$ are defined for all t and n and, for all t, $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$, then X_n converge in law to X.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X - \mu$. Thus it suffices to prove the weak law in the mean zero case.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi'_X(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0}\frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0}\frac{g(\epsilon)}{\epsilon} = 0$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X] = 0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t} e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0) = 0$ and (by chain rule) $g'(0) = \lim_{\epsilon \to 0} \frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon \to 0} \frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0) = g'(0) = 0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X] = 0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t} e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0) = 0$ and (by chain rule) $g'(0) = \lim_{\epsilon \to 0} \frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0} \frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0) = g'(0) = 0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0}\frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0}\frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0)=g'(0)=0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.
- By Lévy's continuity theorem, the A_n converge in law to 0 (i.e., to the random variable that is 0 with probability one).

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

Borel-Cantelli lemmas

- **First Borel-Cantelli lemma:** If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \text{ i.o.}) = 0$.

Borel-Cantelli lemmas

- **First Borel-Cantelli lemma:** If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \text{ i.o.}) = 0$.
- **Second Borel-Cantelli lemma:** If A_n are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ implies $P(A_n \text{ i.o.}) = 1$.

Convergence in probability subsequential a.s. convergence

- **Theorem:** $X_n \to X$ in probability if and only if for every subsequence of the X_n there is a further subsequence converging a.s. to X.

Convergence in probability subsequential a.s. convergence

- **Theorem:** $X_n \to X$ in probability if and only if for every subsequence of the X_n there is a further subsequence converging a.s. to X.
- **Main idea of proof:** Consider event E_n that X_n and X differ by ϵ. Do the E_n occur i.o.? Use Borel-Cantelli.

Pairwise independence example

- **Theorem:** Suppose $A_1, A_2, \ldots$ are pairwise independent and $\sum P(A_n) = \infty$, and write $S_n = \sum_{i=1}^n 1_{A_i}$. Then the ratio S_n/ES_n tends a.s. to 1.

Pairwise independence example

- **Theorem:** Suppose $A_1, A_2, \dots$ are pairwise independent and $\sum P(A_n) = \infty$, and write $S_n = \sum_{i=1}^n 1_{A_i}$. Then the ratio S_n/ES_n tends a.s. to 1.
- **Main idea of proof:** First, pairwise independence implies that variances add. Conclude (by checking term by term) that $\mathrm{Var} S_n \leq ES_n$. Then Chebyshev implies

$$P(|S_n - ES_n| > \delta ES_n) \leq \mathrm{Var}(S_n)/(\delta ES_n)^2 \to 0,$$

which gives us convergence in probability.

Pairwise independence example

- **Theorem:** Suppose $A_1, A_2, \dots$ are pairwise independent and $\sum P(A_n) = \infty$, and write $S_n = \sum_{i=1}^n 1_{A_i}$. Then the ratio S_n/ES_n tends a.s. to 1.
- **Main idea of proof:** First, pairwise independence implies that variances add. Conclude (by checking term by term) that $\mathrm{Var} S_n \leq ES_n$. Then Chebyshev implies

  $$P(|S_n - ES_n| > \delta ES_n) \leq \mathrm{Var}(S_n)/(\delta ES_n)^2 \to 0,$$

  which gives us convergence in probability.
- Second, take a smart subsequence. Let $n_k = \inf\{n : ES_n \geq k^2\}$. Use Borel Cantelli to get a.s. convergence along this subsequence. Check that convergence along this subsequence deterministically implies the non-subsequential convergence.

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

Outline

Weak law of large numbers: characteristic function approach

Laws of large numbers: Borel-Cantelli applications

Strong law of large numbers

General strong law of large numbers

- **Theorem (strong law):** If $X_1, X_2, \dots$ are i.i.d. real-valued random variables with expectation m and $A_n := n^{-1} \sum_{i=1}^{n} X_i$ are the *empirical means* then $\lim_{n\to\infty} A_n = m$ almost surely.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.
- Key to proof is to bound fourth moments of A_n.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.
- Key to proof is to bound fourth moments of A_n.
- $E[A_n^4] = n^{-4}E[S_n^4] = n^{-4}E[(X_1 + X_2 + \ldots + X_n)^4]$.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.
- Key to proof is to bound fourth moments of A_n.
- $E[A_n^4] = n^{-4}E[S_n^4] = n^{-4}E[(X_1 + X_2 + \ldots + X_n)^4]$.
- Expand $(X_1 + \ldots + X_n)^4$. Five kinds of terms: $X_iX_jX_kX_l$ and $X_iX_jX_k^2$ and $X_iX_j^3$ and $X_i^2X_j^2$ and X_i^4.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.
- Key to proof is to bound fourth moments of A_n.
- $E[A_n^4] = n^{-4}E[S_n^4] = n^{-4}E[(X_1 + X_2 + \ldots + X_n)^4]$.
- Expand $(X_1 + \ldots + X_n)^4$. Five kinds of terms: $X_iX_jX_kX_l$ and $X_iX_jX_k^2$ and $X_iX_j^3$ and $X_i^2X_j^2$ and X_i^4.
- The first three terms all have expectation zero. There are $\binom{n}{2}$ of the fourth type and n of the last type, each equal to at most K. So $E[A_n^4] \leq n^{-4}\Big(6\binom{n}{2} + n\Big)K$.

Proof of strong law assuming $E[X^4] < \infty$

- Assume $K := E[X^4] < \infty$. Not necessary, but simplifies proof.
- Note: $\mathrm{Var}[X^2] = E[X^4] - E[X^2]^2 \geq 0$, so $E[X^2]^2 \leq K$.
- The strong law holds for i.i.d. copies of X if and only if it holds for i.i.d. copies of $X - \mu$ where μ is a constant.
- So we may as well assume $E[X] = 0$.
- Key to proof is to bound fourth moments of A_n.
- $E[A_n^4] = n^{-4}E[S_n^4] = n^{-4}E[(X_1 + X_2 + \ldots + X_n)^4]$.
- Expand $(X_1 + \ldots + X_n)^4$. Five kinds of terms: $X_iX_jX_kX_l$ and $X_iX_jX_k^2$ and $X_iX_j^3$ and $X_i^2X_j^2$ and X_i^4.
- The first three terms all have expectation zero. There are $\binom{n}{2}$ of the fourth type and n of the last type, each equal to at most K. So $E[A_n^4] \leq n^{-4}\Big(6\binom{n}{2} + n\Big)K$.
- Thus $E[\sum_{n=1}^\infty A_n^4] = \sum_{n=1}^\infty E[A_n^4] < \infty$. So $\sum_{n=1}^\infty A_n^4 < \infty$ (and hence $A_n \to 0$) with probability 1.

General proof of strong law

- Suppose X_k are i.i.d. with finite mean. Let $Y_k = X_k 1_{|X_k| \leq k}$. Write $T_n = Y_1 + \ldots + Y_n$. **Claim:** $X_k = Y_k$ all but finitely often a.s. so suffices to show $T_n/n \to \mu$. (Borel Cantelli, expectation of positive r.v. is area between cdf and line $y = 1$)

General proof of strong law

- Suppose X_k are i.i.d. with finite mean. Let $Y_k = X_k 1_{|X_k| \leq k}$. Write $T_n = Y_1 + \ldots + Y_n$. **Claim:** $X_k = Y_k$ all but finitely often a.s. so suffices to show $T_n/n \to \mu$. (Borel Cantelli, expectation of positive r.v. is area between cdf and line $y = 1$)
- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to prove it?

General proof of strong law

- Suppose X_k are i.i.d. with finite mean. Let $Y_k = X_k 1_{|X_k| \leq k}$. Write $T_n = Y_1 + \ldots + Y_n$. **Claim:** $X_k = Y_k$ all but finitely often a.s. so suffices to show $T_n/n \to \mu$. (Borel Cantelli, expectation of positive r.v. is area between cdf and line $y = 1$)
- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to prove it?
- **Observe:** $\mathrm{Var}(Y_k) \leq E(Y_k^2) = \int_0^\infty 2yP(|Y_k| > y)dy \leq \int_0^k 2yP(|X_1| > y)dy$. Use Fubini (interchange sum/integral, since everything positive)

$$\sum_{k=1}^{\infty} E(Y_k^2)/k^2 \leq \sum_{k=1}^{\infty} k^{-2} \int_0^\infty 1_{(y<k)} 2yP(|X_1| > y)dy =$$

$$\int_0^\infty \left(\sum_{k=1}^{\infty} k^{-2} 1_{(y<k)}\right) 2yP(|X_1| > y)dy.$$

Since $E|X_1| = \int_0^\infty P(|X_1| > y)dy$, complete proof of claim by showing that if $y \geq 0$ then $2y \sum_{k>y} k^{-2} \leq 4$.

General proof of strong law

- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to use it?

General proof of strong law

- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to use it?
- Consider subsequence $k(n) = [\alpha^n]$ for arbitrary $\alpha > 1$. Using Chebyshev, if $\epsilon > 0$ then

$$\sum_{n=1}^{\infty} P\big(|T_{k(n)} - ET_{k(n)}| > \epsilon k(n)\big) \leq \epsilon^{-1} \sum_{n=1}^{\infty} \mathrm{Var}(T_{k(n)})/k(n)^2$$

$$= \epsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) = \epsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n:k(n)\geq m} k(n)^{-2}.$$

General proof of strong law

- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \le 4E|X_1| < \infty$. How to use it?
- Consider subsequence $k(n) = [\alpha^n]$ for arbitrary $\alpha > 1$. Using Chebyshev, if $\epsilon > 0$ then

$$\sum_{n=1}^{\infty} P\big(|T_{k(n)} - ET_{k(n)}| > \epsilon k(n)\big) \le \epsilon^{-1} \sum_{n=1}^{\infty} \mathrm{Var}(T_{k(n)})/k(n)^2$$

$$= \epsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) = \epsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n:k(n) \ge m} k(n)^{-2}.$$

- **Sum series:**
$\sum_{n:\alpha^n \ge m} [\alpha^n]^{-2} \le 4 \sum_{n:\alpha^n \ge m} \alpha^{-2n} \le 4(1-\alpha^{-2})^{-1} m^{-2}$.

General proof of strong law

- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to use it?
- Consider subsequence $k(n) = [\alpha^n]$ for arbitrary $\alpha > 1$. Using Chebyshev, if $\epsilon > 0$ then

$$\sum_{n=1}^{\infty} P\big(|T_{k(n)} - ET_{k(n)}| > \epsilon k(n)\big) \leq \epsilon^{-1} \sum_{n=1}^{\infty} \mathrm{Var}(T_{k(n)})/k(n)^2$$

$$= \epsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) = \epsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n:k(n)\geq m} k(n)^{-2}.$$

- **Sum series:**
$\sum_{n:\alpha^n \geq m} [\alpha^n]^{-2} \leq 4 \sum_{n:\alpha^n \geq m} \alpha^{-2n} \leq 4(1-\alpha^{-2})^{-1} m^{-2}$.
- **Combine computations (observe RHS below is finite):**

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_k(n)| > \epsilon k(n)) \leq 4(1-\alpha^{-2})^{-1}\epsilon^{-2} \sum_{m=1}^{\infty} E(Y_m^2) m^{-2}.$$

General proof of strong law

- **Claim:** $\sum_{k=1}^{\infty} \mathrm{Var}(Y_k)/k^2 \leq 4E|X_1| < \infty$. How to use it?
- Consider subsequence $k(n) = [\alpha^n]$ for arbitrary $\alpha > 1$. Using Chebyshev, if $\epsilon > 0$ then

$$\sum_{n=1}^{\infty} P\big(|T_{k(n)} - ET_{k(n)}| > \epsilon k(n)\big) \leq \epsilon^{-1} \sum_{n=1}^{\infty} \mathrm{Var}(T_{k(n)})/k(n)^2$$

$$= \epsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) = \epsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n:k(n)\geq m} k(n)^{-2}.$$

- **Sum series:**
  $\sum_{n:\alpha^n \geq m} [\alpha^n]^{-2} \leq 4 \sum_{n:\alpha^n \geq m} \alpha^{-2n} \leq 4(1-\alpha^{-2})^{-1} m^{-2}$.
- **Combine computations (observe RHS below is finite):**

$$\sum_{n=1}^{\infty} P(|T_{k(n)} - ET_k(n)| > \epsilon k(n)) \leq 4(1-\alpha^{-2})^{-1} \epsilon^{-2} \sum_{m=1}^{\infty} E(Y_m^2) m^{-2}.$$

- Since ϵ is arbitrary, get $(T_{k(n)} - ET_{k(n)})/k(n) \to 0$ a.s.

General proof of strong law

- Conclude by taking $\alpha \to 1$. This finishes the case that the X_1 are a.s. positive.

General proof of strong law

- Conclude by taking $\alpha \to 1$. This finishes the case that the X_1 are a.s. positive.
- Can extend to the case that X_1 is a.s. positive within infinite mean.

General proof of strong law

- Conclude by taking $\alpha \to 1$. This finishes the case that the X_1 are a.s. positive.
- Can extend to the case that X_1 is a.s. positive within infinite mean.
- Generally, can consider X_1^+ and X_1^-, and it is enough if one of them has a finite mean.