18.600: Lecture 27

Weak law of large numbers

Scott Sheffield

MIT

Outline

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be a random variable taking only non-negative values. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be a random variable taking only non-negative values. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be a random variable taking only non-negative values. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be a random variable taking only non-negative values. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

- **Proof:** Note that $(X - \mu)^2$ is a non-negative random variable and $P\{|X - \mu| \geq k\} = P\{(X - \mu)^2 \geq k^2\}$. Now apply Markov's inequality with $a = k^2$.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).
- **Markov:** if $E[X]$ is small, then it is not too likely that X is large.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be a random variable taking only non-negative values with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).
- **Markov:** if $E[X]$ is small, then it is not too likely that X is large.
- **Chebyshev:** if $\sigma^2 = \mathrm{Var}[X]$ is small, then it is not too likely that X is far from its mean.

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.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.
- By Chebyshev $P\{|A_n - \mu| \geq \epsilon\} \leq \frac{\mathrm{Var}[A_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.
- By Chebyshev $P\{|A_n - \mu| \geq \epsilon\} \leq \frac{\mathrm{Var}[A_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.
- No matter how small ϵ is, RHS will tend to zero as n gets large.

Outline

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

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?
- Recall that in this strange 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?
- Recall that in this strange 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?
- Recall that in this strange 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 in this case $E[|X|]$ was infinite. Does the weak law hold 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?
- Recall that in this strange 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 in this case $E[|X|]$ was infinite. Does the weak law hold as long as $E[|X|]$ is finite, so that μ is well defined?
- Yes. Can prove this using 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 a random variable and X_n a sequence of random variables.

Continuity theorems

- Let X be a random variable and 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 a random variable and 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 a random variable and 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 (see Wikipedia):** 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 a random variable and 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 (see Wikipedia):** 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 the weak law of large numbers by showing $\lim_{n\to\infty} \phi_{A_n}(t) = \phi_\mu(t) = e^{it\mu}$ for all t. In the special case that $\mu = 0$, this amounts to showing $\lim_{n\to\infty} \phi_{A_n}(t) = 1$ for all t.

Proof of weak law of large numbers in 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 of weak law of large numbers in 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 of weak law of large numbers in 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 of weak law of large numbers in 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 of weak law of large numbers in 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 of weak law of large numbers in 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 of weak law of large numbers in 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).