18.175: Lecture 8

Weak laws and moment-generating/characteristic functions

Scott Sheffield

MIT

Outline

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.
- When X is continuous, can write $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$. So $M(t)$ is a weighted average of a continuum of exponential functions.
- We always have $M(0) = 1$.
- If $b > 0$ and $t > 0$ then $E[e^{tX}] \geq E[e^{t \min\{X,b\}}] \geq P\{X \geq b\} e^{tb}$.
- If X takes both positive and negative values with positive probability then $M(t)$ grows at least exponentially fast in $|t|$ as $|t| \to \infty$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.
- So $M''(0) = E[X^2]$. Same argument gives that nth derivative of M at zero is $E[X^n]$.
- Interesting: knowing all of the derivatives of M at a single point tells you the moments $E[X^k]$ for all integer $k \geq 0$.
- Another way to think of this: write $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \ldots$.
- Taking expectations gives $E[e^{tX}] = 1 + tm_1 + \frac{t^2m_2}{2!} + \frac{t^3m_3}{3!} + \ldots$, where m_k is the kth moment. The kth derivative at zero is m_k.

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.
- Write the moment generating functions as $M_X(t) = E[e^{tX}]$ and $M_Y(t) = E[e^{tY}]$ and $M_Z(t) = E[e^{tZ}]$.
- If you knew M_X and M_Y, could you compute M_Z?
- By independence, $M_Z(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$ for all t.
- In other words, adding independent random variables corresponds to multiplying moment generating functions.

Moment generating functions for sums of i.i.d. random variables

- We showed that if $Z = X + Y$ and X and Y are independent, then $M_Z(t) = M_X(t)M_Y(t)$
- If $X_1 \ldots X_n$ are i.i.d. copies of X and $Z = X_1 + \ldots + X_n$ then what is M_Z?
- Answer: M_X^n. Follows by repeatedly applying formula above.
- This a big reason for studying moment generating functions. It helps us understand what happens when we sum up a lot of independent copies of the same random variable.

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.
- If $Z = X + b$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{tX+bt}] = e^{bt}M_X(t)$.
- Latter answer is the special case of $M_Z(t) = M_X(t)M_Y(t)$ where Y is the constant random variable b.

Existence issues

- Seems that unless $f_X(x)$ decays superexponentially as x tends to infinity, we won't have $M_X(t)$ defined for all t.
- What is M_X if X is standard Cauchy, so that $f_X(x) = \frac{1}{\pi(1+x^2)}$.
- Answer: $M_X(0) = 1$ (as is true for any X) but otherwise $M_X(t)$ is infinite for all $t \neq 0$.
- Informal statement: moment generating functions are not defined for distributions with fat tails.

Outline

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Moment generating functions

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 non-negative random variable. Fix $a > 0$. 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 non-negative random variable 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.
- 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 $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- 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.

L^2 weak law of large numbers

- Say X_i and X_j are uncorrelated if $E(X_iX_j) = EX_iEX_j$.
- Chebyshev/Markov argument works whenever variables are uncorrelated (does not actually require independence).

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.
- Toss αn bins into n balls. How many bins are filled?
- When n is large, the number of balls in the first bin is approximately a Poisson random variable with expectation α.
- Probability first bin contains no ball is $(1 - 1/n)^{\alpha n} \approx e^{-\alpha}$.
- We can explicitly compute variance of the number of bins with no balls. Allows us to show that fraction of bins with no balls concentrates about its expectation, which is $e^{-\alpha}$.

How do you extend to random variables without variance?

- Assume X_n are i.i.d. non-negative instances of random variable X with finite mean. Can one prove law of large numbers for these?
- Try truncating. Fix large N and write $A = X1_{X>N}$ and $B = X1_{X\leq N}$ so that $X = A + B$. Choose N so that EB is very small. Law of large numbers holds for A.

Outline

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Moment generating functions

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?
- 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.
- 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.
- 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.
- 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).

18.175 Theory of Probability
Spring 2014