18.440: Lecture 27

Moment generating functions and characteristic functions

Scott Sheffield

MIT

Outline

Moment generating functions

Characteristic functions

Continuity theorems and perspective

Outline

Moment generating functions

Characteristic functions

Continuity theorems and perspective

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 $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \dots$.
- Taking expectations gives $E[e^{tX}] = 1 + tm_1 + \frac{t^2m_2}{2!} + \frac{t^3m_3}{3!} + \dots$, 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.

Examples

- Let's try some examples. What is $M_X(t) = E[e^{tX}]$ when X is binomial with parameters (p, n)? Hint: try the $n = 1$ case first.
- Answer: if $n = 1$ then $M_X(t) = E[e^{tX}] = pe^t + (1-p)e^0$. In general $M_X(t) = (pe^t + 1 - p)^n$.
- What if X is Poisson with parameter $\lambda > 0$?
- Answer: $M_X(t) = E[e^{tx}] = \sum_{n=0}^{\infty} \frac{e^{tn}e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda}e^{\lambda e^t} = \exp[\lambda(e^t - 1)]$.
- We know that if you add independent Poisson random variables with parameters λ_1 and λ_2 you get a Poisson random variable of parameter $\lambda_1 + \lambda_2$. How is this fact manifested in the moment generating function?

More examples: normal random variables

- What if X is normal with mean zero, variance one?
- $M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\frac{(x-t)^2}{2} + \frac{t^2}{2}\} dx = e^{t^2/2}$.
- What does that tell us about sums of i.i.d. copies of X?
- If Z is sum of n i.i.d. copies of X then $M_Z(t) = e^{nt^2/2}$.
- What is M_Z if Z is normal with mean μ and variance σ^2?
- Answer: Z has same law as $\sigma X + \mu$, so $M_Z(t) = M(\sigma t) e^{\mu t} = \exp\{\frac{\sigma^2 t^2}{2} + \mu t\}$.

More examples: exponential random variables

- What if X is exponential with parameter $\lambda > 0$?
- $M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{-(\lambda-t)x}dx = \frac{\lambda}{\lambda-t}$.
- What if Z is a Γ distribution with parameters $\lambda > 0$ and $n > 0$?
- Then Z has the law of a sum of n independent copies of X. So $M_Z(t) = M_X(t)^n = \left(\frac{\lambda}{\lambda-t}\right)^n$.
- Exponential calculation above works for $t < \lambda$. What happens when $t > \lambda$? Or as t approaches λ from below?
- $M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{-(\lambda-t)x}dx = \infty$ if $t \geq \lambda$.

More examples: 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

Characteristic functions

Continuity theorems and perspective

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 $M_{X+Y} = M_X M_Y$.
- 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 a distinct advantage: they are always well defined for all t even if f_X decays slowly.

Outline

Moment generating functions

Characteristic functions

Continuity theorems and perspective

Perspective

- In later lectures, we will see that one can use moment generating functions and/or characteristic functions to prove the so-called *weak law of large numbers* and *central limit theorem*.
- Proofs using characteristic functions apply in more generality, but they require you to remember how to exponentiate imaginary numbers.
- Moment generating functions are central to so-called *large deviation theory* and play a fundamental role in statistical physics, among other things.
- Characteristic functions are *Fourier transforms* of the corresponding distribution density functions and encode "periodicity" patterns. For example, if X is integer valued, $\phi_X(t) = E[e^{itX}]$ will be 1 whenever t is a multiple of 2π.

Continuity theorems

- Let X be a random variable and X_n a sequence of random variables.
- We say that 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.
- **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.
- **Moment generating analog:** if moment generating functions $M_{X_n}(t)$ are defined for all t and n and $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all t, then X_n converge in law to X.