MATH 320: Probability

Lecture 12: Moment Generating Functions

Chapters 2 and 3: Distributions (2.3, 2.4, 2.6, 2.7, 3.1, 3.2, 3.3)

Moments

Moments

- Definition: The $\boldsymbol{n^{th}}$ **moment** of X is $E(X^n)$.

  Technically, these moments are "about zero (the origin)": $E[(X-0)^n]$.

  The first and second moments are simply $E(X)$ & $E(X^2)$ $\Rightarrow$ V(x) = Second moment - first moment^2

Note ☆

$E(X^3) \neq (E(X))^3$

↳ new RV

- Example: Calculate the third moment of $X \sim \text{Binomial}(n=3, p=0.2)$ using the pmf table below:

x	0	1	2	3
$f(x)$	0.512	0.384	0.096	0.008

$E(X^3) = \sum x^3 f(x) = 0^3(0.512) + 1^3(0.384) + 2^3(0.096) + 3^3(0.008)$

$E(g(x))$

$= 1.368$

- There are other "types" of moments as well.
  - Given random variable X and constant b,

    $E(X-b)$ is the **first moment of $\boldsymbol{X}$ about $\boldsymbol{b}$.**

    $E[(X-b)^n]$ is the $\boldsymbol{n^{th}}$ **moment of $\boldsymbol{X}$ about $\boldsymbol{b}$.**

  - If we let $b = \mu = E(X)$, then we get what are called **central moments**: $E[(X-\mu)^n] \rightarrow$ "about the center $E(X)$".

- Definition: The variance of a random variable X is its second central moment.

  if $n=2 \rightarrow E[(X-\mu)^2] = V(X)$

Moment generating functions

Defining moment generating function and its properties

- The moment generating function (mgf) of random variable X (or the distribution of X), denoted $M_X(t)$, is defined by

Discrete continuous

$$M_X(t) = E(\mathrm{e}^{tX}) = \sum_{x \in \mathcal{X}} e^{tx} f(x) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

- Notes: $E[g(x)]$ where $g(x) = e^{tx}$

  * e^{tX} is a function of t and X and it is a random variable.
  * $E(\mathrm{e}^{tX})$ is the expectation with respect to X, which takes away the randomness of $X \implies$ Result is only a function of t (variable, not constant).

- The mgf has a number of useful properties:

  (1) Finding moments: The derivatives of $M_X(t)$ can be used to find the moments of the random variable X (i.e. take the derivative and evaluate at $t = 0$).

  $$M'_X(0) = E(X), \quad M''_X(0) = E(X^2), \quad \ldots \quad , \quad M_X^{(n)}(0) = E(X^n)$$

  (2) Distribution of a function of a random variable: The mgf of $aX + b$ can be found easily if the mgf of X is known.

  $$M_{aX+b}(t) = \mathrm{e}^{tb} M_X(at)$$

  (3) Uniqueness: If a random variable X has the mgf of a known distribution, then X has that distribution.

- Formalizing this new notation from (1):

$$\boxed{E(X^n) = M_X^{(n)}(0) = \frac{\mathrm{d}}{\mathrm{d}t^n} M_X(t)\bigg|_{t=0}}$$

In words this means: The n^{th} moment of X is equal to the n^{th} derivative of $M_X(t)$ evaluated at $t = 0$.

Derivation / proof / intuition

(1) Finding moments:

Now we can show exactly how / why mgfs generate moments (hence the name).

(a) We will start by looking at the infinite series representation of $\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ and then substituting the random variable tX for x in this series, we obtain:

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots$$

(b) Now we take the expected value of each side. (with respect to x)

$$\underbrace{E(e^{tx})}_{\hookrightarrow = M_X(t)} = 1 + t\,E(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3 E(x^3)}{3!} + \cdots$$

Using the distributive property of expectation and derivation

(c) Then we take the derivative. (with respect to t)

$$M_X'(t) = 0 + E(x) + t\,E(x^2) + \frac{t^2}{2}E(x^3) + \cdots$$

(d) Then we evaluate at $t = 0$.

$$M_X'(0) = 0 + E(x) + \underbrace{0\,E(x^2) + \frac{0^2}{2}E(x^3) + \cdots}_{\text{all have "t" term} \Rightarrow \text{cancel out}}$$

$$= E(x) \rightarrow \text{first moment}$$

(e) Now take the second derivative and evaluate at $t = 0$ again.

$$\rightarrow M_X''(t) = \frac{d}{dt}M_X'(t)$$

$$= 0 + 0 + E(x^2) + t\,E(x^3) + \frac{t^2}{2!}E(x^4) + \cdots$$

$$\rightarrow M_X''(0) = 0 + 0 + E(x^2) + \underbrace{\cancel{0\,E(x^3)} + \cancel{\frac{0^2}{2!}E(x^4)}}_{\text{cancel out}} + \cdots$$

$$= E(X^2) \rightarrow \text{Second moment}$$

(2) Distribution of a function of a random variable:

- Maybe the most important use of mgfs is to find out the distribution of a function of a random variable, specifically $Y = aX + b$.
- Earlier theorems only gave us only $E(aX + b)$ and $V(aX + b)$, but mgfs allow us to find out the distribution, and thus the pmf / cdf.
- Proof of theorem: Let $Y = aX + b$

$$\rightarrow M_Y(t) = E(e^{tY}) \Longleftrightarrow M_{aX+b}(t) = E[e^{(aX+b)t}]$$

$$= E[e^{atx + bt}]$$

$$= E_x[e^{atx}\,\underline{e^{bt}}]$$

constant value with respect to X (for the expectation)

$$= e^{bt}E[e^{atx}]$$

$$= e^{bt}M_X(at)$$

(3) Uniqueness:

- When we introduced pmfs / pdfs and cdfs of the commonly used distributions, we stated a useful property:

  If we recognize that we have a pmf / pdf where the the range of the random variable and the probabilities / density function match the scenario of a specific distribution, then that random variable must follow that specific distribution.

- This is true for mgfs as well.

  Mgfs are unique. This means that if a random variable X has the moment during function of a known random variable, it must be that kind of variable.

- We will not prove this.

Example: Using mgfs

$\sim Bin(n=3,\ p=0.2)$

(a) Find the moment generating function of X, $M_X(t)$.

	x	0	1	2	3
$g(x)$	e^{tx}	$e^{t(0)} = 1$	e^{t}	e^{2t}	e^{3t}
	$f(x)$	0.512	0.384	0.096	0.008

$M_X(t) = E(e^{tX}) = \sum e^{tx} f(x)$

$\rightarrow M_X(t) = 0.512(1) + 0.384e^{t} + 0.096e^{2t} + 0.008e^{3t}$

(b) Find $E(X)$ using the mgf found in part (a).

Step 1: Take the derivative of $M_X(t)$ with respect to t.

$\rightarrow M_X'(t) = \frac{d}{dt} M_X(t)$

$= 0 + 0.384e^{t} + 0.096(2)e^{2t} + 0.008(3)e^{3t}$

Step 2: Now evaluate the derivative at $t = 0$ (i.e. plug in $t = 0$ and simplify).

$\rightarrow M_X'(0) = M_X'(t)\Big|_{t=0}$

$= 0 + 0.384\cancel{e^{0}} + 0.096(2)\cancel{e^{2(0)}} + 0.008(3)\cancel{e^{3(0)}}$

$= 0.6 \overset{\checkmark}{=} E(X)$

Confirm
Binomial mean $= np = 3(0.2) \overset{\checkmark}{=} 0.6$

Calc review e^{x}

$\rightarrow \frac{d}{dx} e^{x} = e^{x}$

$\rightarrow \frac{d}{dx} e^{f(x)} = f'(x)\, e^{f(x)}$

$\rightarrow \frac{d}{dx} e^{ax} = a e^{ax}$

(c) Find $V(X)$ using the mgf.

Note that the higher derivatives can be used in the same way $\Longrightarrow$ Take the second derivative of $M_X(t)$ with respect to t and then evaluate at $t = 0$.

$\rightarrow M_X''(t) = \frac{d}{dt}\left[M_X'(t)\right]$

$= 0 + 0.384e^{t} + 0.096(2^2)e^{2t} + 0.008(3^2)e^{3t}$

$\rightarrow M_X''(0) = M_X''(t)\Big|_{t=0}$

$= 0 + 0.384\cancel{e^{0}} + 0.096(2^2)\cancel{e^{2(0)}} + 0.008(3^2)\cancel{e^{3(0)}}$

$= 0.84 = E(X^2)$

Confirm
variance $\rightarrow V(X) = E(X^2) - (E(X))^2 \Longrightarrow E(X^2) = V(X) + (E(X))^2 = npq + (np)^2$ (For Binomial)

$= 3(0.2)(0.8) + (3(0.2))^2 \overset{\checkmark}{=} 0.84$

Final points

- Expanding on property (1) of mgfs (for discrete random variables, replace summation with integration for continuous).

$$M_X(t) = E(e^{tX}) = \sum e^{tx} f(x)$$

Middle step: $M_X'(t) =$ $\sum x e^{tx} f(x)$ $\qquad$ $M_X''(t) =$ $\sum x^2 e^{tx} f(x)$

Property: $M_X'(0) =$ $\sum x f(x) = E(x)$ $\qquad$ $M_X''(0) =$ $\sum x^2 f(x) = E(x^2)$

- The general form is the following:

$$\boxed{\begin{aligned} M_X^{(n)}(t) &= \sum x^n e^{tx} f(x) \\ M_X^{(n)}(0) &= \sum x^n f(x) = E(X^n) \end{aligned}}$$

- Variance: Using these new ideas and notations, we have a new way to write the variance using mgfs:

$$\boxed{V(x) = \quad M_X''(0) - \left[M_X'(0)\right]^2 = M_X''(t)\big|_{t=0} - \left[M_X'(t)\big|_{t=0}\right]^2}$$

- Mgf's for random variables don't always exist. There is technically more to the definition $M_X(t) = E(e^{tX})$.
  - Condition: It must be the case that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E(e^{tX})$ exists. If the expectation does not exist in the neighborhood of 0, we say that the mgf does not exist.
  - Said another way: Derivations of $M_X(t)$ of all orders exist at $t = 0 \Longrightarrow M_X(t)$ is continuous at $t = 0$.
- Many standard probability distributions have moment generating functions which can be found fairly easily. This gives us another way to derive the mean and variance formulas stated previously.

Mgfs of commonly used discrete distributions

Discrete uniform random variable mgf

- The mgf for a discrete uniform random variable is straightforward to find.
- Let $X \sim$ Discrete uniform (N_0, N_1)

$M_X(t) = E(e^{tX}) = \sum_{x=N_0}^{N_1} e^{tx} \underbrace{f(x)}_{= \frac{1}{N_1 - N_0 + 1}}$

$= \frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} e^{tx}$

- If $X \sim$ Discrete uniform (N_0, N_1)

$$M_X(t) = \frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} \mathrm{e}^{tx}$$

Bernoulli random variable mgf

- The mgf for a bernoulli random variable is straightforward to find.
- Let $X \sim$ Bernoulli (p)

x	0	1
e^{tx}	$\mathrm{e}^{t0} = 1$	e^{t}
$f(x)$	$1 - p = q$	p

$M_X(t) = 1q + pe^t$

- If $X \sim$ Bernoulli (p)

$$M_X(t) = (1 - p) + p\mathrm{e}^t = q + p\mathrm{e}^t$$

Binomial random variable mgf

- The mgf for a binomial random variable is easy to find in the simplest cases. Then we can generalize the pattern without proof.

  1. We just saw the mgf when $X \sim \text{Binomial}\,(n=1,p) \sim$ Bernoulli (p)

     $M_X(t) = (q + pe^t)^{1}$

  2. Now let $X \sim \text{Binomial}\,(n=2,p)$.

x	0	1	2
e^{tx}	1	e^t	e^{2t}
$f(x)$	q^2	$2pq$	p^2

$M_X(t) = q^2 + 2pqe^t + p^2e^{2t}$

$= \underbrace{q^2}_{a} + \underbrace{2}_{2}\underbrace{q}_{a}\underbrace{(pe^t)}_{b} + \underbrace{(pe^t)^2}_{b^2}$

$= (q + pe^t)^2$

$\rightarrow$ Now extrend pattern for n trials

- If $X \sim \text{Binomial}\,(n,p)$

$$\boxed{M_X(t) = (q + pe^t)^n}$$

- To derive the mean and variance of the binomial distribution using the mgf, we would simply need to do the following:
  - $E(X)$: Take the derivative of $M_X(t)$ and evaluate at $t = 0$.
  - $V(X)$: Take the second derivative of $M_X(t)$ and evaluate at $t = 0$.

    Then use the result of $E(X)$ and the alternate variance form $V(X) = E(X^2) - [E(X)]^2$.
  - Warning: This requires careful attention when taking the derivatives in order to correctly keep track of everything.

- Example: You are working with a random variable X, and find that it's mgf is:

  $M_X(t) = (0.2 - 0.8e^t)^7 \overset{}{\underset{\text{uniqueness}}{\Longrightarrow}} X \sim \text{Binomial}(n=7, p=0.8)$

Geometric random variable mgf

- The derivation of the mgf for a geometric random variable is not bad either. It relies on the sum of an infinite geometric series.

- Let $X \sim \text{Geometric}\,(p)$ $\rightarrow \mathcal{X} = \{1, 2, 3, \ldots\}$

$\rightarrow M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} \underbrace{f(x)}_{= q^{x-1}p}$

$= \frac{p}{q} \sum_{x=1}^{\infty} (\underbrace{qe^t}_{a})^x$

$= \frac{p}{q}\left[\frac{qe^t}{1 - qe^t}\right]$

$= \frac{pe^t}{1 - qe^t}$

Geometric series

$\sum_{x=1}^{\infty} a^x = \frac{a}{1-a}$

Restriction is to ensure mgf is not undefined
(i.e. divide by zero)

- If $X \sim \text{Geometric}\,(p)$

$$M_X(t) = \frac{pe^t}{1 - qe^t} \qquad t < -\ln(q)$$

- Just like with the binomial, need to be careful when deriving the mean and variance using the mgf.
- Example: Let $X \sim \text{Geometric}\,(p)$. Suppose $Y = X - 1$. Find the mgf of Y, $M_Y(t)$.

Transformation

$$M_Y(t) = M_{X-1}(t) = e^{-t} M_X(t)$$

$$= \cancel{e^{-t}} \frac{p\cancel{e^t}}{1 - qe^t}$$

$$= \frac{p}{1 - qe^t}$$

Negative binomial random variable mgf

- We will not derive this. But we can make use of the pattern / relationship of Bernoulli and binomial.

$$\sum_{i=1}^{n} \text{Bernoulli} = \text{Binomial} \qquad \sum_{i=1}^{r} \text{Geometric} = \text{Negative Binomial}$$

$$M_X(t) \to q + pe^t \to (q + pe^t)^n \qquad \frac{pe^t}{1 - qe^t} \to \left[\frac{pe^t}{1 - qe^t}\right]^r$$

- If $X \sim \text{Negative binomial}\,(r, p)$

$$M_X(t) = \left[\frac{pe^t}{1 - qe^t}\right]^r \qquad t < -\ln(q)$$

Poisson random variable mgf

- The derivation of the mgf for a Poisson random variable is quite short as well, but will be left for grad school. It makes use of the series for e^x.
- If $X \sim \text{Poisson}\,(\lambda)$

$$M_X(t) = e^{\lambda(e^t - 1)}$$

- Derivation: Use the mgf to derive the mean and variance of a Poisson(λ) random variable.

(1) $E(X) = M_X'(0)$

$\to M_X'(t) = \frac{d}{dt} M_X(t) = \frac{d}{dt}\left[e^{\lambda(e^t-1)}\right]$

$\downarrow \quad = \lambda e^t e^{\lambda(e^t-1)}$

$\to M_X'(0) = \lambda \cancel{e^0} \cancel{e^{\lambda(e^0-1)}}$ $\;=1 \;\; =0$

$\downarrow \; \overset{\checkmark}{=} \lambda$

Calc Review: Product rule

Let $u = f(x)$ & $v = g(x)$

$(uv)' = u'v + uv'$

(2) $\to E(X^2) = M_X''(0)$

$\to M_X''(t) = \frac{d}{dt} M_X'(t) = \frac{d}{dt}\left[\underbrace{\lambda e^t}_{u}\ \underbrace{e^{\lambda(e^t-1)}}_{v}\right]$

$\downarrow \quad = \lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)}$

$\to M_X''(0) = \lambda e^0 \cancel{e^{\lambda(e^0-1)}} + (\lambda \cancel{e^0})^2 \cancel{e^{\lambda(e^0-1)}}$

$\downarrow \quad = \lambda + \lambda^2$

(3) $\to V(X) = E(X^2) - [E(X)]^2$

$= \lambda + \lambda^2 - (\lambda)^2$

$\downarrow \; \overset{\checkmark}{=} \lambda$

- Example: Let $X \sim \text{Poisson}\,(\lambda = 2)$. Suppose $Y = 3X + 5$. Find the mgf of Y, $M_Y(t)$.

Transformation

$M_Y(t) = M_{3X+5}(t) = e^{5t} M_X(3t)$

$\downarrow \quad = e^{5t}\left[e^{2(e^{3t}-1)}\right]$

Hypergeometric random variable mgf

- Apparently it exists, but we will not discuss it.

Moment generating functions for continuous random variables

Inroduction

- Some continuous random variables have useful mgfs, which can be written in closed form and easily applied, and others do not.

  We will discuss the mgf of the uniform, gamma and normal distributions. The beta and lognormal distributions do not have useful mgfs, and the Pareto mgf does not exist.

- Note that we could find mgfs of any of the non-named distribution examples we have seen before. They would just require application of the definition and probably lots of integration by parts (as most $f(x)$'s were non-simple functions of X, e.g. ROI example $f(x) = 0.75(1 - x^2), \quad -1 \le x \le 1$).

  But the mgfs we will discuss here are much more interesting and common.

Uniform mgf

- The derivation of the uniform mgf is just a direct application of the definition.

$$M_X(t) = E(e^{tX}) = \int_a^b e^{tx} \frac{1}{b-a}\, dx$$

$$= \frac{1}{t(b-a)} e^{tx} \Big| \; f(b) - f(a)$$

$$= \frac{e^{tb} - e^{ta}}{t(b-a)}$$

- If $X \sim \text{Uniform}\,(a, b)$,
$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}, \quad t \neq 0$$

Gamma mgf

- The gamma mgf can be easily derived. It requires the identity which was shown earlier when introducing the gamma function $\Gamma(\alpha)$:

$$\int_0^\infty x^{\alpha-1}\,\mathrm{e}^{-\beta x}\,\mathrm{d}x = \frac{\Gamma(\alpha)}{\beta^\alpha}, \qquad x>0 \quad \text{and} \quad \alpha,\beta>0.$$

- To derive $M_X(t)$ for $X \sim \text{Gamma}\,(\alpha,\beta)$ we just need to go from the definition. This will also be left for grad school.

- If $X \sim \text{Gamma}\,(\alpha,\beta)$,

$$M_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

- If we wanted to, we could then easily go through the process of taking the derivatives of $M_X(t)$ and evaluating at $t=0$ to show that

$$E(X) = \frac{\alpha}{\beta}, \qquad\qquad V(X) = E(X^2) - [E(X)]^2 = \frac{\alpha}{\beta^2}$$

Exponential mgf

- Since the exponential distribution is a special case of the gamma distribution,

$X \sim \text{Gamma}\,(\alpha=1, \beta) = \text{Exp}(\beta)$

we also Know [redacted] the mgf of the exponential distribution.

- If $X \sim \text{Exponential}\,(\beta)$,

$$M_X(t) = \frac{\beta}{\beta - t}, \quad t < \beta$$

Normal mgf

- The normal mgf can actually be derived without fancy calculus skills, just fancy algebra skills.

- If $X \sim \text{Normal}\,(\mu,\sigma^2)$,

$$M_X(t) = \mathrm{e}^{\mu t + \frac{\sigma^2 t^2}{2}}$$

- Lets derive the mean of the normal distribution using the mgf:

$$\to M_X'(t) = \frac{d}{dt}\left[e^{\mu t + \frac{\sigma^2 t^2}{2}}\right]$$

$$\downarrow \quad = (\mu + \sigma^2 t)\left[e^{\mu t + \frac{\sigma^2 t^2}{2}}\right]$$

$$\to M_X'(0) = (\mu + \sigma^2(0))\, \underbrace{\cancel{e^{\mu(0) + \frac{\sigma^2(0^2)}{2}}}}_{=1}$$

$$\downarrow \quad = \mu \overset{\checkmark}{=} E(X)$$

- Derivation of the variance is straightforward, just requires careful application of the product rule.

  Question: What must $E(X^2)$ equal?

  $E(x^2) = V(x) + (E(x))^2$

  $= \sigma^2 + \mu^2$

- We can now prove the following theorem that was shown earlier without proof:

  Linear transformation of normal random variables: If $X \sim \text{Normal}\,(\mu, \sigma^2)$ and $Y = aX + b$. Then

$$Y \sim \text{Normal}\,(a\mu + b, a^2\sigma^2)$$

$M_{ax+b}(t) = e^{tb} M_x(at)$

$= e^{tb} e^{\mu(at) + \frac{\sigma^2 (at)^2}{2}}$

$= e^{(a\mu + b)t + \frac{(a^2\sigma^2) t^2}{2}}$

$\sim$ Normal $(a\mu + b,\ a^2\sigma^2)$ MGF

- Now lets apply the same strategy to the specific linear transformation of standardizing and show the mgf of the standard normal distribution Z.

$\rightarrow Z = \frac{x-\mu}{\sigma} = \frac{1}{\sigma} x - \frac{\mu}{\sigma}$

$\rightarrow M_Z(t) = M_{\frac{1}{\sigma}x - \frac{\mu}{\sigma}}(t) = e^{-\frac{\mu}{\sigma}t} M_X(\frac{1}{\sigma}t) = e^{-\frac{\mu}{\sigma}t} e^{\mu(\frac{1}{\sigma}t) + \frac{\sigma^2(\frac{1}{\sigma}t)^2}{2}}$

$= e^{0t + \frac{t^2}{2}}$

$\sim$ Normal $(\mu = 0,\ \sigma^2 = 1)$

- Lets try this strategy for a distribution other than normal.

  In the previous section, we had an example about standardizing an exponential distribution $X \sim \text{Exponential}\,(\lambda)$. We showed the mean and variance of the standardized exponential:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 1/\lambda}{1/\lambda} = \lambda X - 1 \qquad \Longrightarrow \qquad E(Z) = 0, \qquad V(Z) = 1$$

  But did not have a conclusion about the distribution of Z.

$M_Z(t) = M_{\lambda X - 1}(t) = e^{-t} M_X(\lambda t)$

$= e^{-t}\left[\frac{\lambda}{\lambda - \lambda t}\right]$

$= e^{-t}\left[\frac{\lambda}{\lambda(1-t)}\right]$

$= \frac{1}{e^t(1-t)}$ $\neq$ Exponential Mgf

$\Longrightarrow Z \not\sim \text{Exp}$