Stat 110

Unit 6: Moments

Chapter 6
in the text

Unit 6 Outline

- Summarizing a Distribution
- Moments of a Distribution
- Sample Moments
- Moment Generating Functions (MGFs)
- Sums of Independent R.V.s via MGFs

Summaries of a distribution

- To this point, we've seen two very important (possibly the two most important) summaries of a distribution. What are they?
  - Mean and Variance
- What do they measure?
- The mean is a *measure of central tendency* of a random variable because it tells us something about the center of the distribution
- The variance is a measure of *spread* or *variation* of a random variable.
- There are other measures of center and spread in a distribution. And there are other summaries we may like to consider.

Definitions: Median and Mode

- Two more measures of central tendency:
- We say that c is a ***median*** of a r.v. X if $P(X \leq c) \geq 1/2$ and if $P(X \geq c) \geq 1/2$.
- The simplest way this can happen is if the CDF of X hits 1/2 exactly at c, but we know that some CDFs have jumps.

- For a discrete r.v. X, we say that c is a ***mode*** of X if it maximizes the PMF: $P(X = c) \geq \mathrm{P}(X = x)$ for all x.
  For a continuous r.v. X with PDF f, we say that c is a ***mode*** if it maximizes the PDF: $f(c) \geq f(x)$ for all x.

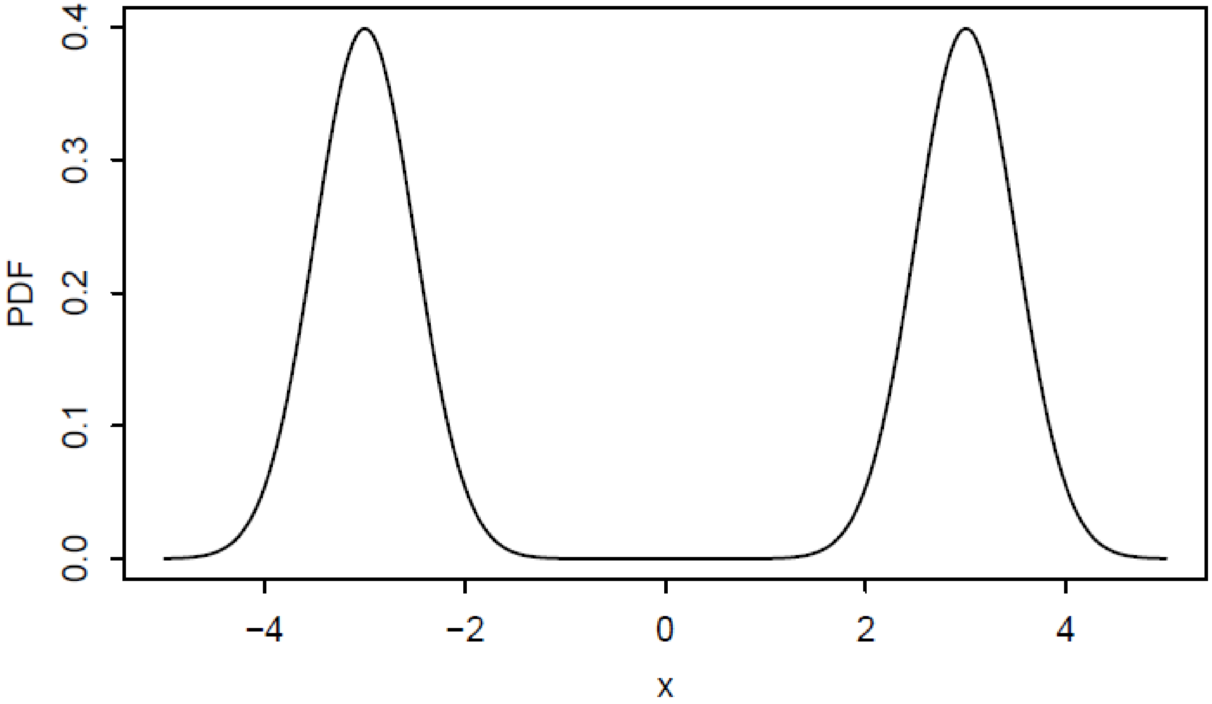
- The median can be thought of as the *midpoint* of the probability in a distribution, and the mode can be thought of as the point of highest probability or density.
- Note: they may not be unique.

Calculating median and mode

- As with the mean, the median and mode of a r.v. depend only on its distribution, so we can talk about the mean, median, or mode of a distribution without referring to any particular r.v. that has that distribution.
- For example, if $Z \sim N(0, 1)$ then the median of Z is 0 (since $\Phi(0) = 1/2$ by symmetry), and we also say that the standard Normal distribution has median 0.
- This also means that if a CDF F is continuous and strictly increasing, then $F^{-1}(1/2)$ is the unique median of a distribution.
- How can we calculate the mode of a continuous distribution?
- Take the derivative of the PDF and set to zero.

Visualizing mean, median, and mode

- It is straightforward to determine the mean, median, and mode of a r.v. based on a plot of its PDF:
- For example: what are the mean, median and mode of this r.v.?



Ex: Discrete mean, median, and mode

- A certain company has100 employees. Let $s_1, s_2, \ldots, s_{100}$ be their salaries, sorted in increasing order (we can still do this even if some salaries appear more than once). Let X be the salary of a randomly selected employee (chosen uniformly). The mean, median, and mode for the data set $s_1, s_2, \ldots, s_{100}$ are defined to be the corresponding quantities for X.
- What are the mean, median, and mode of X?
- What is a typical salary? What is the most useful one-number summary of the salary data? The answer, as is often the case, is it *depends on the goal*.

Two important properties

- Let X be an r.v. with mean μ, and let m be a median of X.
- The value of c that minimizes the mean squared error, $E[(X-c)^2]$, is $c = \mu$.
- A value of c that minimizes the mean absolute error, $E(|X-c|)$, is $c = m$.
- How do we prove/derive these results?
- Second one is easy (do piecewise: if $c >$ m vs. $c <$ m).
- The first one is easy if you first show that for any c:

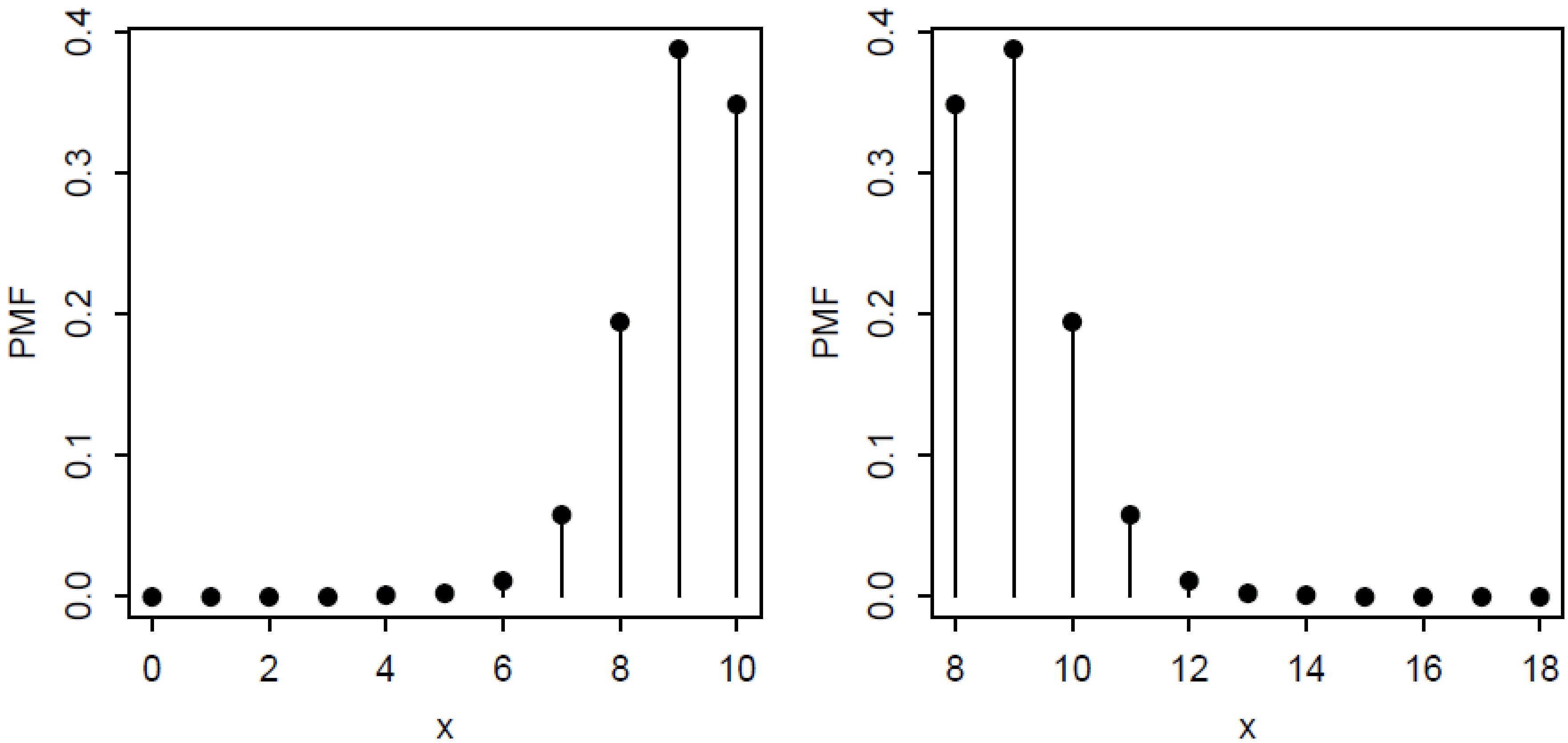
$$E[(X-c)^2] = \mathrm{Var}(X) + (c-\mu)^2$$

Going beyond center and spread

- There are other summary measures of a distribution beyond those for center and spread.
- Why is this important? Because two r.v.s can have very different distributions, but have the same mean and variance.
- For example: let $X \sim \text{Bin}(10, 0.9)$, and let $(Y - 8) \sim \text{Bin}(10, 0.1)$. How do X and Y compare?

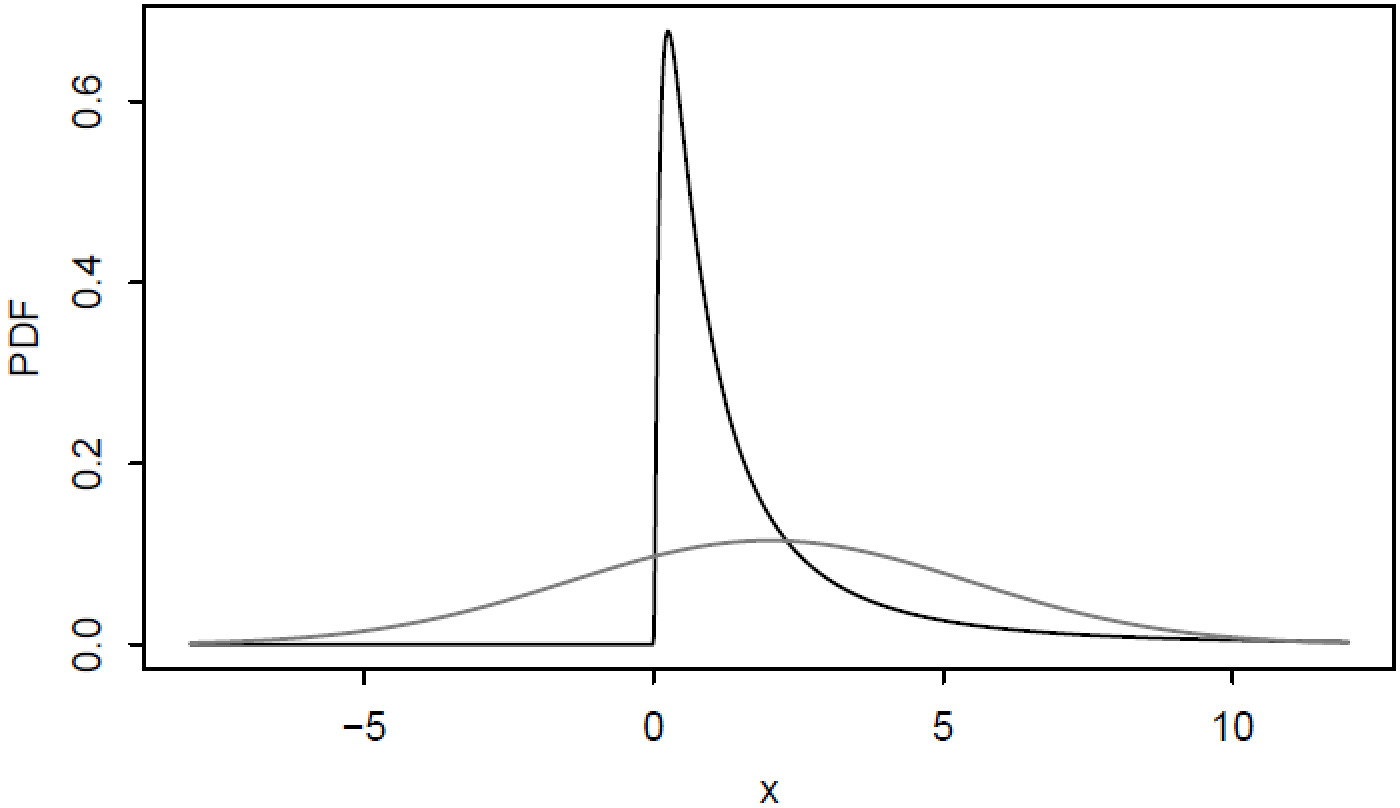
- We can start to talk about *skewness*, etc… of a distribution.
- This leads to more formally describing the moments of a distribution.

$X \sim \text{Bin}(10, 0.9)$ vs. $Y \sim [\text{Bin}(10, 0.1) + 8]$



- We call *X* left-skewed and *Y* right-skewed.
- Skewness follows the longer tail. It affects the mean more than the median.

Continuous Example



- These are PDFs of a $N(2,12)$ r.v. and a *Log-Normal* r.v. with the same mean and variance.

Unit 6 Outline

- Summarizing a Distribution
- Moments of a Distribution
- Sample Moments
- Moment Generating Functions (MGFs)
- Sums of Independent R.V.s via MGFs

Definitions: Moments and Central Moments

- The k^{th} ***moment*** (k is an integer) of a r.v. X is defined as the expectation $E(X^k)$.
- The k^{th} moment is said to exist if $E(|X^k|) < \infty$.

- The k^{th} ***central moment*** (k is an integer) of a r.v. X is defined as the expectation $E[(X\text{-}\mu)^k]$.
- The k^{th} central moment is said to exist…

- What moments are the mean and variance?

Definition: Skewness

- The ***skewness*** of a r.v. X with mean μ and variance σ^2 is the third standardized moment of X:

$$\mathrm{Skew}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

- Why do we standardize?
  - So the result does not depend on the mean μ and variance σ^2. And so it is unit-less.
- What is it measuring?
  - Asymmetry.

Definition: Symmetry

- A r.v. X is said to have a ***symmetric distribution about μ*** if $X - \mu$ has the same distribution as $\mu - X$. This can also be stated as *"X* is symmetric" or "the distribution of X is symmetric".
- In terms of the PDF, this means that $f(x) = f(2\mu - x)$.
- Why does that make sense?
- What does the plot of the PDF for a symmetric r.v. look like?
- What will be the value of *skewness* for a symmetric r.v.?
- What about for any odd numbered central moment?

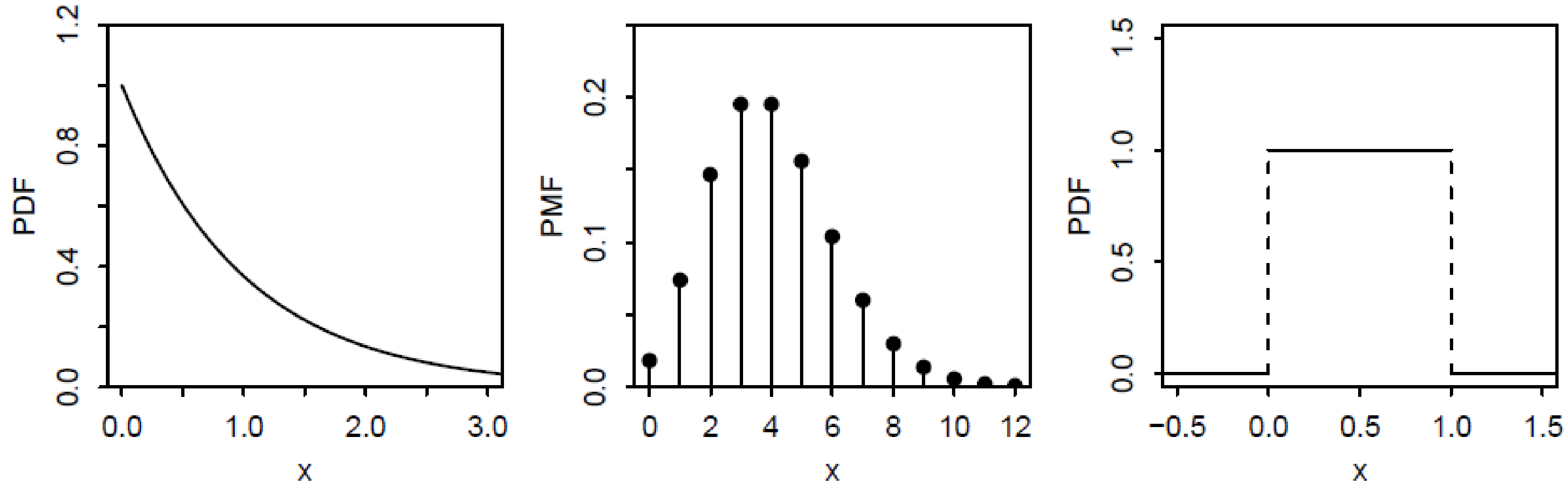
Definition: Kurtosis

- The ***kurtosis*** of a r.v. X with mean μ and variance σ^2 is the fourth standardized moment of X:

$$\mathrm{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

- Why the "-3"? So that any Normal distribution has kurtosis 0.
- What is Kurtosis measuring?
- It is measuring how *heavy tailed* a distribution is.
- A prototypical distribution with large kurtosis has a PDF with a sharp peak in the center (within 1σ), low shoulders ($\pm$(1 to 2)σ), and heavy tails (beyond $\pm 2\sigma$).

Skewness and Kurtosis Examples



- Skewness and kurtosis of some named distributions:
  Left: Expo(1) PDF, skewness = 2, kurtosis = 6.
  Middle: Pois(4) PMF, skewness = 0.5, kurtosis = 0.25.
  Right: Unif(0, 1) PDF, skewness = 0, kurtosis = -1.2.

Unit 6 Outline

- Summarizing a Distribution
- Moments of a Distribution
- Sample Moments
- Moment Generating Functions (MGFs)
- Sums of Independent R.V.s via MGFs

Definition: Sample Moments

- Let $X_1,\ldots,X_n$ be i.i.d. random variables. The k^{th} ***sample moment*** is the r.v.

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

- The ***sample mean*** $\overline{X}_n$ is the first sample moment:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- In contrast, the ***population mean*** or true mean is $\mu = E(X_i)$, the mean of the distribution from which the X_i were drawn.

Mean and Variance of $\bar{X}_n$

- Let $X_1,\ldots,X_n$ be i.i.d. random variables with mean μ and variance σ^2. Then $\bar{X}_n$ is unbiased for estimating μ. That is:

$$E\left(\bar{X}_n\right) = \mu$$

- And the variance of $\bar{X}_n$ is:

$$\mathrm{Var}\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$$

- Proof:

Sample Variance

- Let $X_1,\ldots,X_n$ be i.i.d. random variables with mean and variance . Then the ***sample variance*** is the r.v.:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- And the ***sample standard deviation*** is the square root of the sample variance.

- Why the (n-1)? Because that way it is unbiased for estimating σ^2.
- Proof:

- Note: this is more of a Stat 111 kinda thing.

Unit 6 Outline

- Summarizing a Distribution
- Moments of a Distribution
- Sample Moments
- Moment Generating Functions (MGFs)
- Sums of Independent R.V.s via MGFs

Generating Moments

- We just learned that moments can be used to summarize a distribution.
- It certainly would be nice to have some function that we could use to generate them.
- And in fact that is something called a Moment Generating Function (MGF) to do just that.
- Not only that, but a MGF is unique to a distribution. So if we can determine a r.v. has a specific MGF, then we know exactly what distribution it has!

Definition: Moment Generating Functions (MGFs)

- The ***moment generating function*** (MGF) of a r.v. X is $M(t) = E(e^{tX})$, as a function of t, if this is finite on some open interval $(-a, a)$ containing 0. Otherwise, the MGF of X does not exist.

$$M(t) = E(e^{tX})$$

- So what is an MGF? As the name suggests, it is a function that encodes the *moments* of a distribution.

MGF Example: Geometric and Uniform

- Let $X \sim \text{Geom}(p)$. Calculate the MGF for X.

$$M(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} q^k p = p \sum_{k=0}^{\infty} \left(qe^t\right)^k = \frac{p}{1 - qe^t}$$

- For what values of t?
  For t in $(-\infty, \ln(1/q))$, an open interval containing 0.

- Let $U \sim \text{Unif}(a,b)$. Calculate the MGF for U.

$$M(t) = E(e^{tX}) = \int_a^b e^{tu} \left[\frac{1}{b-a}\right] du = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

- For $t \neq 0$. Why should $M(0) = 1$?

Calculating Moments from MGFs

- Given the MGF of X, we can get the n^{th} moment of X by evaluating the n^{th} derivative of the MGF at 0:
  $E(X^n) = M^{(n)}(0)$.
- Wow, that makes life simple!
- For example, we can find $E(X)$ for the Geometric(p) dist.:

$$M(t) = \frac{p}{1-qe^t} \Rightarrow M'(t) = \frac{pqe^t}{\left(1-qe^t\right)^2}$$

$$\Rightarrow E(X) = M'(0) = \frac{pqe^0}{\left(1-qe^0\right)^2} = \frac{pq}{(1-q)^2} = \frac{q}{p}$$

- In general, why does this work out this way?
- Let's look at the Taylor series expansion of $M(t)$ about 0, and of $E(e^{tX})$.

Calculating Moments from MGFs

- Let's look at the Taylor series expansion of $M(t)$ about 0:

$$M(t) = M(0) + M'(0)t + M''(0)\frac{t^2}{2} + M'''(0)\frac{t^3}{3!} = \sum_{n=0}^{\infty} M^{(n)}(0)\frac{t^n}{n!}$$

- But we also have:

$$M(t) = E(e^{tX}) = E\left(\sum_{n=0}^{\infty} X^n \frac{t^n}{n!}\right) = \sum_{n=0}^{\infty} E\left(X^n\right)\frac{t^n}{n!}$$

- And by matching terms (it's ok to do this since we said $E(e^{tX})$ is finite in an interval around 0), we can show that:

$$E(X^n) = M^{(n)}(0)$$

MGFs determine the distribution

- The MGF of a random variable determines its distribution: if two r.v.s have the same MGF, they must have the same distribution. In fact, if there is even a tiny interval ($-a$, a) containing 0 on which the MGFs are equal, then the r.v.s must have the same distribution.
- The above theorem is a difficult result in analysis, so we will not prove it here.
- But the result is useful! Again, that is:
- **If two r.v.s have the same MGF, they must have the same distribution.**

Unit 6 Outline

- Summarizing a Distribution
- Moments of a Distribution
- Sample Moments
- Moment Generating Functions (MGFs)
- Sums of Independent R.V.s via MGFs

MGF of a sum of independent r.v.s

- If X and Y are independent, then the MGF of $X + Y$ is the product of the individual MGFs:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

- This is true because if X and Y are independent, then:

$$E(e^{t(X+Y)}) = E(e^{t(X)}e^{t(Y)}) = E(e^{t(X)})E(e^{t(Y)})$$

- Using this fact, we can get the MGFs of the Binomial and Negative Binomial, which are sums of independent Bernoullis and Geometrics, respectively.

MGF of a sum of independent r.v.s

- What is the MGF of a Bern(p) r.v.?

$$M(t) = E(e^{tX}) = e^{t0}q + e^{t1}p = q + pe^t$$

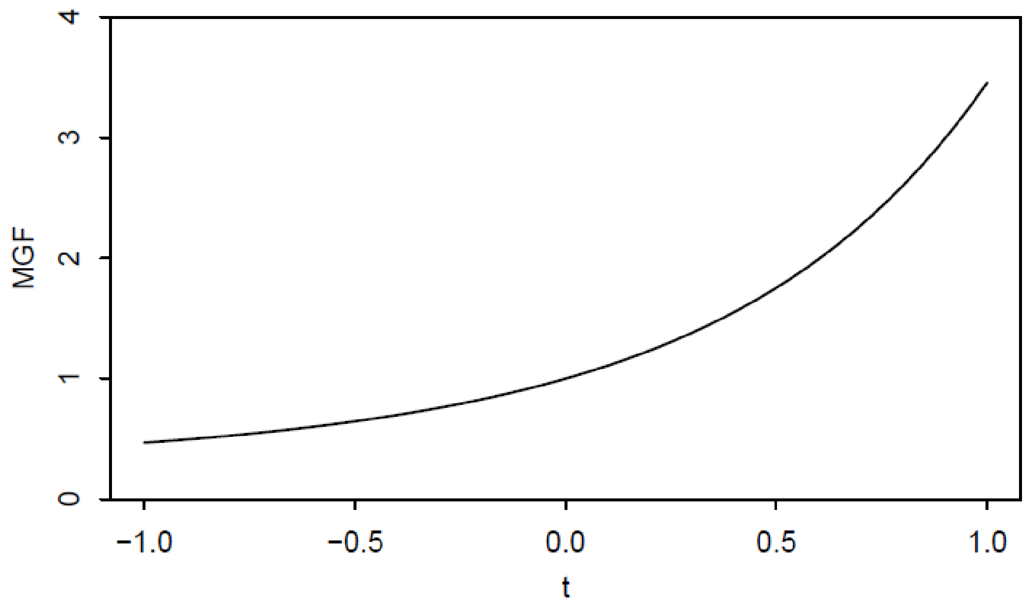
- Use the result from the previous page to calculate the MGF of a Bin(p) r.v.

$$M(t) = \left(q + pe^t\right)^n$$

- Use the result from the previous page and the MGF of a Geom(p) r.v. to calculate the MGF of a NBin(r, p) r.v.

$$M(t) = \left(\frac{p}{1-qe^t}\right)^r$$

Visualizing an MGF as a function



- The MGF of $X \sim \text{Bin}(2, 0.5)$, $M(t) = (0.5e^t + 0.5)^2$.
- How are the mean and second moment depicted here?
- The slope of the MGF at $t = 0$ is 1, so the mean of the distribution is 1. The concavity or second derivative of the MGF at $t = 0$ is 3/2, so the second moment of the distribution is 3/2.

Warning: not all r.v.s have an MGF

- Not all r.v.s have an MGF. Some r.v.s X don't even have $E(X)$ exist, or don't have $E(X^k)$ exist for some $k > 1$, in which case the MGF clearly will not exist.
- But even if all the moments of X exist, the MGF may not exist if the moments grow too quickly. Luckily, there is a way to fix this: inserting an imaginary number!
- The function $\psi(t) = E(e^{itX})$ with $i = \sqrt{-1}$ is called the *characteristic function* by statisticians and the *Fourier transform* by everyone else. It turns out that the characteristic function always exists.
- In this class we will focus on the MGF rather than the characteristic function, to avoid having to handle imaginary numbers. Using $\psi(t)$ is more of a Stat 210 idea.

Effect of location-scale transformation on an MGF

- If X has MGF $M(t)$, then the MGF of $a + bX$ is:

$$E(e^{t(a+bX)}) = e^{ta}E(e^{btX}) = e^{at}M(bt)$$

- Let's use this to calculate the MGF of the Normal and Exponential distributions.

- So first calculate the MGF for a standard Normal, and then apply the above result to generalize it.
- Same thing for an Exponential distribution.

MGF of a Normal r.v.

- First, the MGF of $Z \sim N(0,1)$ is:

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

- After completing the square (and recognizing the PDF of a $N(t,1)$ r.v., we get:

$$M_Z(t) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dz = e^{t^2/2}$$

- Then the MGF of $X \sim N(\mu,\sigma^2)$ is:

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} = e^{\mu t + (1/2)\sigma^2 t^2}$$

MGF of an Exponential r.v.

- First, the MGF of X ~ Expo(1) is (for $t < 1$):

$$M_X(t) = E(e^{tZ}) = \int_0^\infty e^{tx} e^{-x} dx$$

$$= \int_0^\infty e^{-(1-t)x} dx = \frac{1}{1-t}$$

- Then the MGF of $Y = X/\lambda$ ~ Expo(λ) is (for $t < \lambda$):

$$M_Y(t) = M_X(t/\lambda) = \frac{1}{1-t/\lambda} = \frac{\lambda}{\lambda - t}$$

Generating All the Moments

- As explained earlier, we can calculate moments by taking derivatives of $M(t)$ and setting $t = 0$. But for some distributions, there's an easier way to create ALL the moments.
- If we can match the MGF to a known series, and then match it up with a Taylor series approximation, then getting ALL the moments is easy!
- For example, let's look at the MGF for an Expo(1) r.v.:

$$M(t) = \frac{1}{1-t}$$

- What series does this look like the result of?
  - A geometric series!

Generating All the Moments

- So the MGF for an Expo(1) r.v. is equivalent to:

$$M(t) = \frac{1}{1-t} = \sum_{n=1}^{\infty} t^n$$

- And how do we make this look like a Taylor series expansion?

$$M(t) = \frac{1}{1-t} = \sum_{n=1}^{\infty} t^n = \sum_{n=1}^{\infty} n! \frac{t^n}{n!}$$

- And we already know that: $M(t) = \sum_{n=1}^{\infty} E(X^n) \frac{t^n}{n!}$
- Thus: $E(X^n) = n!$ for all n for this distribution. Wow!
- This same approach can be used for the Normal and Log-Normal Distributions as well.

Log-Normal Distribution

- A r.v. Y is said to have a *Log-Normal* distribution with parameters μ and σ^2 if $Y = e^X$ where $X \sim N(\mu,\sigma^2)$. We denote this as $Y \sim LN(\mu,\sigma^2)$ or $Y \sim LogN(\mu,\sigma^2)$.
- We will determine the PDF for the Log-Normal distribution in an upcoming unit (unit 8).
- We mention it now because it does not have a defined MGF since $E(e^{tY})$ is infinite for all $t > 0$.
- Consider the case when $Y = e^Z$ for $Z \sim N(0,1)$; by LOTUS:

$$E(e^{tY}) = E(e^{te^Z}) = \int_{-\infty}^{\infty} e^{te^z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{te^z - z^2/2} dz$$

- But $te^z - z^2/2$ goes to infinity as z grows. Thus we say that the MGF of Y does not exist.

MGF of a sum of independent r.v.s

- Back to the point: we now have a new strategy to determine the distribution of a sum of r.v.'s:
- Multiply the individual MGFs together to see if the product is recognizable as the MGF of a named distribution.
- Example: Sum of independent Normal Distributions.
- Let $X_1 \sim N(\mu_1,\sigma_1^2)$ and independently $X_2 \sim N(\mu_2,\sigma_2^2)$.
- Find the MGF of $(X_1 + X_2)$. What named distribution's MGF does this match?

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t+(1/2)\sigma_1^2 t^2} e^{\mu_2 t+(1/2)\sigma_2^2 t^2}$$

$$= e^{(\mu_2+\mu_2)t+(1/2)(\sigma_1^2+\sigma_2^2)t^2}$$

- Thus, a sum of independent Normal r.v.s is Normal itself! This agrees 100% with the Normal dist.'s story ☺

Take Home Message

- Distributions can be summarized based on various measures
  - Means, Medians, Modes for location/center
  - Variance & Stand Deviation for spread
  - Skewness (asymmetry) and Kurtosis (tailedness) for shape
- Moments are a specific way to summarize distributions
  - There are standard moments: $E(X^k)$
  - And central moments: $E[(X-\mu)^k]$
- Sample moments (ex: $\bar{X}$) are important r.v.s (especially in Stat 111)
- Moments Generating Function (MGFs) are a useful tool:
  - $M_X(t) = E(e^{tX})$
  - MGFs can be used to generate the moments based on taking derivatives at $t = 0$. $E(X) = M_X'(0)$, $E(X^2) = M_X''(0)$,
  - MGFs are unique to distributions, so if two r.v.s have the same MGF, then they have the same distribution ☺

Last Word

- You can think of MGFs like cookbooks:
  - Cookbooks contain the information (recipes) to create various dishes
  - MGFs contain the information (derivatives at $t = 0$) to create the various Moments of a distribution

$$\begin{tikzcd} & \boxed{M_X(t)} \arrow[dl] \arrow[d] \arrow[dr] & \\ \boxed{E(X)} & \boxed{E(X^2)} \quad \ldots & \boxed{E(X^k)} \end{tikzcd}$$