11: Moment-Generating Functions

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 20, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability
- Independence of random variables, and covariance/correlation
- Sums of Random Variables; Indicators

Moment Generating Functions

Leadup

- Suppose we have two random variables X and Y.
- If $\mathbb{E}(X) = \mathbb{E}(Y)$, can we conclude that X and Y have the same distribution (sometimes notated $X \stackrel{\mathrm{d}}{=} Y$)?
    - No! Counterexample: $X \sim \mathrm{Bin}(20, 0.1)$ and $Y \sim \mathrm{Pois}(2)$.
- What if, in addition to $\mathbb{E}(X) = \mathbb{E}(Y)$, we have $\mathrm{Var}(X) = \mathrm{Var}(Y)$?
    - Still No! Counterexample: $X \sim \mathrm{Geom}(0.5)$ and $Y \sim \mathrm{Pois}(2)$.

- So, what *is* enough?
- Turns out, equality in *all* moments is enough; $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ for every $n \in \mathbb{N}$.
- That's a lot of moments we need to check! Wouldn't it be nice if there is some quantity that gives us access to the moments of a distribution?

MGF

- There *is* such a quantity, and it is called the **Moment Generating Function**.

Definition: Moment Generating Function

The **Moment Generating Function** of X, denoted $M_X(t)$, is defined as

$$M_X(t) := \mathbb{E}\left[e^{Xt}\right] \tag{1}$$

- As it stands, this definition works equally well for discrete and continuous random variables! Now, it is true that exactly *how* we compute the expectation on the RHS depends on whether X is discrete or continuous; specifically,

$$M_X(t) = \begin{cases} \displaystyle\sum_k e^{kt} p_X(k) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{\mathbb{R}} e^{xt} f_X(x) \, \mathrm{d}x & \text{if } X \text{ is continuous} \end{cases}$$

- Why the name? Because of the following theorem:

Theorem

Given a random variable X with moment-generating function $M_X(t)$, we have that

$$\mathbb{E}[X^n] = M_X^{(n)}(0)$$

provided that $M_X(t)$ is finite in an interval containing the origin. Here, $M_X^{(n)}$ denotes the n^{th} derivative of M_X.

- I may post a proof for this in a bit, for those who are curious.

Example

- Suppose $X \sim \text{Geom}(p)$.
- Then $M_X(t) := \mathbb{E}(e^{Xt}) \sum_k e^{kt}\mathbb{P}(X = k)$

$$= \sum_{k=1}^{\infty} e^{kt} \cdot p \cdot (1-p)^{k-1}$$

$$= \frac{p}{1-p} \sum_{k=1}^{\infty} \left[(1-p)e^t\right]^k$$

$$= \frac{p}{\cancel{1-p}} \times \frac{\cancel{(1-p)}\, e^t}{1-(1-p)e^t} = \frac{pe^t}{1-(1-p)e^t}$$

- Of course, this is valid only if the geometric series above converges, which occurs when $(1-p)e^t < 1 \implies t < -\ln(1-p)$; otherwise, the MGF is infinite. Thus,

$$\boxed{M_X(t) = \begin{cases} \frac{pe^t}{1-(1-p)e^t} & \text{if } t < -\ln(1-p) \\ \infty & \text{otherwise} \end{cases}}$$

Example

- With this formula, we can re-derive the expectation of the Geometric Distribution. Assuming $t < -\ln(1-p)$, we have

$$
\begin{aligned}
M_X'(t) &= \frac{pe^t \cdot [1-(1-p)e^t] - pe^t \cdot [-(1-p)e^t]}{[1-(1-p)e^t]^2} \\
&= \frac{pe^t \cancel{-p(1-p)e^{2t}} + \cancel{p(1-p)e^{2t}}}{[1-(1-p)e^t]^2} \\
&= \frac{pe^t}{[1-(1-p)e^t]^2}
\end{aligned}
$$

$$
M_X'(0) = \frac{p \cdot e^0}{[1-(1-p)e^0]^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}
$$

Example

Suppose $X \sim \text{Exp}(\lambda)$.

(a) Derive an expression for $M_X(t)$, the moment-generating function (MGF) of X. Be sure to specify where the MGF is finite and where it is infinite!

(b) Use your answer to part (a) to derive a formula for $\mathbb{E}[X^n]$, where $n \in \mathbb{N}$.

Some Common MGF's of Discrete Distributions

Distribution	MGF at t
$\text{Bin}(n, p)$	$(1 - p + pe^t)^n, \quad \forall t \in \mathbb{R}$
$\text{Geom}(p)$	$\begin{cases} \dfrac{pe^t}{1-(1-p)e^t} & \text{if } t < -\ln(1-p) \\ \infty & \text{otherwise} \end{cases}$
$\text{NegBin}(r, p)$	$\begin{cases} \left(\dfrac{pe^t}{1-(1-p)e^t}\right)^r & \text{if } t < -\ln(1-p) \\ \infty & \text{otherwise} \end{cases}$
$\text{Pois}(\lambda)$	$e^{\lambda(e^t-1)}, \quad \forall t \in \mathbb{R}$

Some Common MGF's of Continuous Distributions

Distribution	MGF at t
$\text{Exp}(\lambda)$	$\begin{cases} \dfrac{\lambda}{\lambda - t} & \text{if } t < \lambda \\ 0 & \text{otherwise} \end{cases}$
$\text{Gamma}(r, \lambda)$	$\begin{cases} \left(\dfrac{\lambda}{\lambda - t}\right)^r & \text{if } t < \lambda \\ 0 & \text{otherwise} \end{cases}$
$\mathcal{N}(\mu, \sigma^2)$	$\exp\left\{\mu t + \frac{\sigma^2}{2} \cdot t^2\right\}; \quad \forall t \in \mathbb{R}$
$\text{Unif}[a, b]$	$\begin{cases} \dfrac{e^{tb} - e^{ta}}{t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$

Equality in Distribution

- Let me go back to one of the points I made at the beginning of this lecture; namely, that MGF's are enough to determine a distribution.
- I'll phrase this a bit more formally:

Theorem

Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$, respectively. Suppose there exists a $\delta > 0$ such that for every $t \in (-\delta, \delta)$ we have $M_X(t) = M_Y(t)$ [and that both of these values are finite]. Then X and Y have the same distribution.

- A slight rephrasing:

Theorem

Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$, respectively. If $M_X(t) = M_Y(t)$ for all t, then X and Y have the same distribution [i.e. the same pmf's/pdf's]

- So, for example, suppose X is a random variable with MGF

$$M_X(t) = \begin{cases} \frac{0.2e^t}{1-0.8e^t} & \text{if } t < -\ln(0.8) \\ \infty & \text{otherwise} \end{cases}$$

Then, we can immediately conclude that $X \sim \text{Geom}(0.2)$, since the MGF is continuous and finite over a small interval containing the origin.

MGF's of Transformations

Theorem

Given a random variable X with MGF $M_X(t)$, and another random variable $Y := aX + b$ for constants a, b, then $M_Y(t) = e^{bt} M_X(at)$.

Proof.

- By the definition of MGF's,
$$M_Y(t) := \mathbb{E}[e^{tY}]$$
- Since $Y = aX + b$, we can substitute $aX + b$ in place of Y in our equation above:

$$M_Y(t) = \mathbb{E}[e^{t(aX+b)}] = \mathbb{E}[e^{taX+tb}] = \mathbb{E}[e^{(at)X} e^{bt}] = e^{bt}\mathbb{E}[e^{(at)X}] = e^{bt} M_X(at)$$

□

Example

Suppose X is a random variable with MGF given by

$$M_X(t) = \begin{cases} \frac{0.2e^{3t}}{1-0.8e^{3t}} & \text{if } t < -1/3 \cdot \ln(0.8) \\ \infty & \text{otherwise} \end{cases}$$

and say I wish to compute $\mathbb{P}(X = 3)$. Here is the logic:

- The MGF looks a bit like that of the Geom(0.2) distribution; as such, suppose $Y \sim \text{Geom}(0.2)$. Then

$$M_Y(t) = \begin{cases} \frac{0.2e^{t}}{1-0.8e^{t}} & \text{if } t < -\ln(0.8) \\ \infty & \text{otherwise} \end{cases}$$

- Now, suppose $X = 3Y$. Then, by the previous theorem,

$$M_X(t) = M_Y(3t) \begin{cases} \frac{0.2e^{3t}}{1-0.8e^{3t}} & \text{if } 3t < -\ln(0.8) \\ \infty & \text{otherwise} \end{cases}$$

  which is indeed the MGF we started with.
- Hence, $X = 3Y$ where $Y \sim \text{Geom}(0.2)$, meaning

$$\mathbb{P}(X = 3) = \mathbb{P}(3Y = 3) = \mathbb{P}(Y = 1) = (1 - 0.2)^{1-1} \cdot (0.2) = \boxed{0.2}$$

MGF's of Sums

Theorem

Given two independent random variables X and Y with MGF's $M_X(t)$ and $M_Y(t)$, respectively, and given a new random variable $Z := X + Y$, we have

$$M_Z(t) = M_X(t) \cdot M_Y(t)$$

Proof.

- By the definition of MGF's,
$$M_Z(t) := \mathbb{E}[e^{tZ}]$$
- Since $Z = X + Y$, we can substitute $X + Y$ in place of Z in our equation above:
$$M_Z(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}]$$
- We know that functions of independent random variables are also independent; hence, since $X \perp Y$ we have $e^{tX} \perp e^{tY}$, and so
$$M_Z(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}] = \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = M_X(t) \cdot M_Y(t)$$

□

MGF's of Sums

Theorem

Given a collection of independent random variables X_i each with MGF $M_{X_i}(t)$, and defining $S := \sum_{i=1}^{n} X_i$, we have

$$M_S(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Example

- We have previously seen that if $X, Y \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$, then $(X + Y) \sim \text{Gamma}(2, \lambda)$. The way we proved this before was using the convolution formula.
- We can re-derive this result much quicker using MGF's. Observe:

$$
\begin{aligned}
M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) && \text{[by independence]} \\
&= \left(\begin{cases} \dfrac{\lambda}{\lambda - t} & \text{if } t < \lambda \\ 0 & \text{otherwise} \end{cases} \right) \cdot \left(\begin{cases} \dfrac{\lambda}{\lambda - t} & \text{if } t < \lambda \\ 0 & \text{otherwise} \end{cases} \right) && \text{[MGF of Exp]} \\
&= \begin{cases} \left(\dfrac{\lambda}{\lambda - t} \right)^2 & \text{if } t < \lambda \\ \infty & \text{otherwise} \end{cases} && \text{[MGF of Exp]}
\end{aligned}
$$

  which we recognize as the MGF of the $\text{Gamma}(2, \lambda)$ distribution.
- This can be generalized to derive the sum of n i.i.d. $\text{Exp}(\lambda)$ distributed random variables, or even to derive the distribution of the sum of n independent $\text{Gamma}(r_i, \lambda)$ distributions!

Example/Theorem

Theorem

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ aned $Y \sim \mathcal{N}(\mu_Y,\ \sigma_Y^2)$ with $X \perp Y$, then

$$(X+Y) \sim \mathcal{N}\left(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2\right)$$

Proof.

On the Chalkboard. $\square$

Theorem

If we have a collection of independent random variables $X_i \sim \mathcal{N}(\mu_i,\ \sigma_i^2)$, then

$$\left(\sum_{i=1}^{n} a_i X_i\right) \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Proof.

Omitted. □

Inversions?

- Now, everything we have done thus far (by way of using MGF's to identify distributions) has required us to recognize the MGF that results.
- What happens if that's not the case?
- In other words, given an MGF, is there a way to "invert" the MGF to obtain the original p.m.f./p.d.f., without having to resort to lookup tables?
- The answer, surprisingly, is "not really!"
- There is one exception, however:

Theorem

Given a random variable X with MGF given by

$$M_X(t) = \sum_{i=1}^{n} p_i e^{tk_i}; \quad \forall t \in \mathbb{R}$$

for constants k_i and p_i such that $\sum_{i=1}^{n} p_i = 1$, then the p.m.f. of X is given by $p_X(k_i) = p_i$ for all $i = 1, \cdots, n$.

Example

Suppose X has MGF given by

$$M_X(t) = \frac{1}{5}e^{-4t} + \frac{3}{5} + \frac{1}{5}e^{3.2t}, \quad \forall t \in \mathbb{R}$$

- Note that this MGF is of the form listed in the previous theorem with $n = 3$ and $k_1 = -4,\ k_2 = 0$, and $k_3 = 3.2$ (note that there is a "hidden" e^{0t} attached to the (3/5) in the MGF). This means that the state space of X is

$$S_X = \{-4, 0, 3.2\}$$

- Additionally, the PMF values can be read off directly as the coefficients associated with each of the exponential terms:

$\boldsymbol{k}$	-4	0	3.2
$\boldsymbol{p_X(k)}$	$1/5$	$3/5$	$1/5$

Example

- By the way: now that we have the PMF of X, we can compute $\mathbb{E}[X]$ in two ways.
- Using MGF's:

$$M_X'(t) = -4 \cdot \frac{1}{5} e^{-4t} + 3.2 \cdot \frac{1}{5} e^{3.2t}$$

$$\mathbb{E}[X] = M_X'(0) = -4 \cdot \frac{1}{5} e^{-4 \cdot 0} + 3.2 \cdot \frac{1}{5} e^{3.2 \cdot 0} = -4 \cdot \frac{1}{5} + 3.2 \cdot \frac{1}{5} = \boxed{-0.16}$$

- Using the definition of expectation:

$$\begin{aligned} \mathbb{E}[X] &= \sum_k k p_X(k) \\ &= (-4) \cdot \left(\frac{1}{5}\right) + (0) \cdot \left(\frac{3}{5}\right) + (3.2) \cdot \left(\frac{1}{5}\right) = \boxed{-0.16} \end{aligned}$$