

Chapter 5. Multiple Random Variables

5.6: Moment Generating Functions

[Slides \(Google Drive\)](#)

Alex Tsun

[Video \(YouTube\)](#)

Last time, we talked about how to find the distribution of the sum of two independent random variables. Some of the most important use cases are to prove the results we've been using for so long: the sum of independent Binomials is Binomial, the sum of independent Poissons is Poisson (we proved this in 5.5 using convolution), etc. We'll now talk about Moment Generating Functions, which allow us to do these in a different (and arguably easier) way. These will also be used to prove the Central Limit Theorem (next section), probably the most important result in all of statistics! Also, to derive the Chernoff bound (6.2). The point is, these are used to *prove* a lot of important results. They might not be as direct applicable to problems though.

5.6.1 Moments

First, we need to define what a moment is.

Definition 5.6.1: Moments

Let X be a random variable and $c \in \mathbb{R}$ a scalar. Then: The k^{th} moment of X is:

$$\mathbb{E}[X^k]$$

and the k^{th} moment of X (about c) is:

$$\mathbb{E}[(X - c)^k]$$

The first four moments of a distribution/RV are commonly used, though we have only talked about the first two of them. I'll briefly explain each but we won't talk about the latter two much.

1. The first moment of X is the mean of the distribution $\mu = \mathbb{E}[X]$. This describes the center or average value.
2. The second moment of X about μ is the variance of the distribution $\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mu)^2]$. This describes the spread of a distribution (how much it varies).
3. The third standardized moment is called skewness $\mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]$ and typically tells us about the asymmetry of a distribution about its peak. If skewness is positive, then the mean is larger than the median and there are a lot of extreme high values. If skewness is negative, then the median is larger than the mean and there are a lot of extreme low values.
4. The fourth standardized moment is called kurtosis $\mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}[X^4]}{\sigma^4}$, which measures how peaked a distribution is. If the kurtosis is positive, then the distribution is thin and pointy, and if the kurtosis is negative, the distribution is flat and wide.

5.6.2 Moment Generating Functions (MGFs)

We'll first define the MGF of a random variable X , and then explain its use cases and importance.

Definition 5.6.2: Moment Generating Functions (MGFs)

Let X be a random variable. The **moment generating function (MGF)** of X is a function of a dummy variable t :

$$M_X(t) = \mathbb{E} [e^{tX}]$$

If X is discrete, by LOTUS:

$$M_X(t) = \sum_{x \in \Omega_X} e^{tx} p_X(x)$$

If X is continuous, by LOTUS:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

We say that the MGF of X exists, if there is a $\varepsilon > 0$ such that the MGF is finite for all $t \in (-\varepsilon, \varepsilon)$, since it is possible that the sum or integral diverges.

Let's do some example computations before discussing why it might be useful.

Example(s)

Find the MGF of the following random variables:

(a) X is a discrete random variable with PMF:

$$p_X(k) = \begin{cases} 1/3 & k = 1 \\ 2/3 & k = 2 \end{cases}$$

(b) Y is a Unif(0, 1) continuous random variable.

Solution

(a)

$$\begin{aligned} M_X(t) &= \mathbb{E} [e^{tX}] \\ &= \sum_x e^{tx} p_X(x) && \text{[LOTUS]} \\ &= \frac{1}{3}e^t + \frac{2}{3}e^{2t} \end{aligned}$$

(b)

$$\begin{aligned}
M_Y(t) &= \mathbb{E} [e^{tY}] \\
&= \int_0^1 e^{ty} f_Y(y) dy && \text{[LOTUS]} \\
&= \int_0^1 e^{ty} \cdot 1 dy && [f_Y(y) = 1, 0 \leq y \leq 1] \\
&= \frac{e^t - 1}{t}
\end{aligned}$$

□

5.6.3 Properties and Uniqueness of MGFs

There are some useful properties of MGFs that we will discuss. Let X, Y be *independent* random variables, and $a, b \in \mathbb{R}$ be scalars. Then, recall that the moment generating function of X is: $M_X(t) = \mathbb{E} [e^{tX}]$.

1. Computing MGF of Linear Transformations: We'll first see how we can compute the MGF of $aX + b$ if we know the MGF of X :

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

2. Computing MGF of Sums: We can also compute the MGF of the sum of independent RVs X and Y given their individual MGFs: (the third step is due to independence):

$$M_{X+Y}(t) = \mathbb{E} [e^{t(X+Y)}] = \mathbb{E} [e^{tX} e^{tY}] = \mathbb{E} [e^{tX}] \mathbb{E} [e^{tY}] = M_X(t) M_Y(t)$$

3. Generating Moments with MGFs: The reason why MGFs are named they way they are, is because they *generate moments* of X . That means, they can be used to compute $\mathbb{E} [X]$, $\mathbb{E} [X^2]$, $\mathbb{E} [X^3]$, and so on. How? Let's take the derivative of an MGF (with respect to t):

$$M'_X(t) = \frac{d}{dt} \mathbb{E} [e^{tX}] = \frac{d}{dt} \sum_{x \in \Omega_X} e^{tx} p_X(x) = \sum_{x \in \Omega_X} \frac{d}{dt} (e^{tx} p_X(x)) = \sum_{x \in \Omega_X} x e^{tx} p_X(x)$$

note in the last step that x is a constant with respect to t and so $\frac{d}{dt} e^{tx} = x e^{tx}$.

Note that if evaluate the derivative at $t = 0$, we get $\mathbb{E} [X]$ since $e^0 = 1$:

$$M'_X(0) = \sum_{x \in \Omega_X} x e^{0x} p_X(x) = \sum_{x \in \Omega_X} x p_X(x) = \mathbb{E} [X]$$

Now, let's consider the second derivative:

$$M''_X(t) = \frac{d}{dt} M'_X(t) = \frac{d}{dt} \sum_{x \in \Omega_X} x e^{tx} p_X(x) = \sum_{x \in \Omega_X} \frac{d}{dt} (x e^{tx} p_X(x)) = \sum_{x \in \Omega_X} x^2 e^{tx} p_X(x)$$

If we evaluate the second derivative at $t = 0$, we get $\mathbb{E}[X^2]$:

$$M_X''(0) = \sum_{x \in \Omega_X} x^2 e^{0x} p_X(x) = \sum_{x \in \Omega_X} x^2 p_X(x) = \mathbb{E}[X^2]$$

Seems like there's a pattern - if we take the n -th derivative of $M_X(t)$, then we will generate the n -th moment $\mathbb{E}[X^n]$!

Theorem 5.6.1: Properties and Uniqueness of Moment Generating Functions

For a function $f : \mathbb{R} \rightarrow \mathbb{R}$, we will denote $f^{(n)}(x)$ to be the n^{th} derivative of $f(x)$. Let X, Y be independent random variables, and $a, b \in \mathbb{R}$ be scalars. Then MGFs satisfy the following properties:

1. $M_X'(0) = \mathbb{E}[X]$, $M_X''(0) = \mathbb{E}[X^2]$, and in general $M_X^{(n)}(0) = \mathbb{E}[X^n]$. This is why we call M_X a *moment generating* function, as we can use it to generate the moments of X .
2. $M_{aX+b}(t) = e^{tb} M_X(at)$.
3. If $X \perp Y$, then $M_{X+Y}(t) = M_X(t) M_Y(t)$.
4. **(Uniqueness)** The following are equivalent:
 - (a) X and Y have the same distribution.
 - (b) $f_X(z) = f_Y(z)$ for all $z \in \mathbb{R}$.
 - (c) $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$.
 - (d) There is an $\varepsilon > 0$ such that $M_X(t) = M_Y(t)$ for all $t \in (-\varepsilon, \varepsilon)$ (they match on a small interval around $t = 0$).

That is M_X uniquely identifies a distribution, just like PDFs or CDFs do.

We proved the first three properties before stating all the theorems, so all that's left is property 4. This is a very complex proof (out of the scope of this course), but we can prove it for a special case.

Proof of Property 4 for a Special Case. We'll prove that, if X, Y are discrete rvs with range $\Omega = \{0, 1, 2, \dots, m\}$ and whose MGFs are equal everywhere, that $p_X(k) = p_Y(k)$ for all $k \in \Omega$. That is, if two distributions have the same MGF, they have the same distribution (PMF).

For any t , we have

$$M_X(t) = M_Y(t)$$

By definition of MGF, we get

$$\sum_{k=0}^m e^{tk} p_X(k) = \sum_{k=0}^m e^{tk} p_Y(k)$$

Subtracting the right-hand side from both sides gives:

$$\sum_{k=0}^m e^{tk} (p_X(k) - p_Y(k)) = 0$$

Let $a_k = p_X(k) - p_Y(k)$ for $k = 0, \dots, m$ and write e^{tk} as $(e^t)^k$. Then, we get

$$\sum_{k=0}^m a_k (e^t)^k = 0$$

Note that this is an m -th degree polynomial in e^t , and remember that this equation holds for (uncountably) infinitely many t . An m^{th} degree polynomial can only have m roots, unless all the coefficients are 0. Hence $a_k = 0$ for all k , and so $p_X(k) = p_Y(k)$ for all k . \square

Now we'll see how to use MGFs to prove some results we've been using.

Example(s)

Suppose $X \sim \text{Poi}(\lambda)$, meaning X has range $\Omega_X = \{0, 1, 2, \dots\}$ and PMF:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Compute $M_X(t)$.

Solution First, let's recall the Taylor series:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} p_X(k) = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} (e^t)^k \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} && \text{[Taylor series with } x = \lambda e^t \text{]} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

\square

We can use MGFs in our proofs of certain facts about RVs.

Example(s)

If $X \sim \text{Poi}(\lambda)$, compute $\mathbb{E}[X]$ using its MGF we computed earlier $M_X(t) = e^{\lambda(e^t - 1)}$.

Solution We can prove that $\mathbb{E}[X] = \lambda$ as follows.

First we take the derivative of the moment generating function (don't forget the chain rule of calculus) and see that:

$$M'_X(t) = e^{\lambda(e^t - 1)} \cdot \lambda e^t$$

Then, we know that:

$$\mathbb{E}[X] = M'_X(0) = e^{\lambda(e^0 - 1)} \cdot \lambda e^0 = \lambda$$

\square

Example(s)

If $Y \sim \text{Poi}(\gamma)$ and $Z \sim \text{Poi}(\mu)$ and $Y \perp Z$, show that $Y + Z \sim \text{Poi}(\gamma + \mu)$ using the uniqueness property of MGFs. (Recall we did this exact problem using convolution in 5.5).

Solution First note that a $\text{Poi}(\gamma + \mu)$ RV has MGF $e^{(\gamma + \mu)(e^t - 1)}$ (just plugging in $\gamma + \mu$ as the parameter). Since Y and Z are independent, by property 3,

$$M_{Y+Z}(t) = M_Y(t)M_Z(t) = e^{\gamma(e^t - 1)}e^{\mu(e^t - 1)} = e^{(\gamma + \mu)(e^t - 1)}$$

The MGF of $Y + Z$ which we computed is the same as that of $\text{Poi}(\gamma + \mu)$. So, by the uniqueness of MGFs (which implies that an MGF can uniquely describe a distribution), $Y + Z \sim \text{Poi}(\gamma + \mu)$.

Which way was easier for you - this approach or using convolution? MGF's have limitations though whereas convolution doesn't (besides independence) - we need to compute the MGF of Y, Z but we also need to know the MGF of what distribution we are trying to "get". \square

Example(s)

Now, use MGFs to prove the closure properties of Gaussian RVs (which we've been using without proof).

- If $V \sim \mathcal{N}(\mu, \sigma^2)$ and $W \sim \mathcal{N}(\nu, \gamma^2)$ are independent, that $V + W \sim \mathcal{N}(\mu + \nu, \sigma^2 + \gamma^2)$.
- If $a, b \in \mathbb{R}$ are constants and $X \sim \mathcal{N}(\mu, \sigma^2)$, show that $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

You may use the fact that if $Y \sim \mathcal{N}(\mu, \sigma^2)$, that

$$M_Y(t) = \int_{-\infty}^{\infty} e^{ty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Solution

- If $V \sim \mathcal{N}(\mu, \sigma^2)$ and $W \sim \mathcal{N}(\nu, \gamma^2)$ are independent, we have the following:

$$M_{V+W}(t) = M_V(t)M_W(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} e^{\nu t + \frac{\gamma^2 t^2}{2}} = e^{(\mu + \nu)t + \frac{(\sigma^2 + \gamma^2)t^2}{2}}$$

This is the MGF of a Normal distribution with mean $\mu + \nu$ and variance $\sigma^2 + \gamma^2$. So, by uniqueness of MGFs, $Y + Z \sim \mathcal{N}(\mu + \nu, \sigma^2 + \gamma^2)$.

- Let us examine the moment generating function for $aX + b$. (We'll use the notation $\exp(z) = e^z$ so that we can actually see what's in the exponent clearly):

$$M_{aX+b}(t) = e^{bt} M_X(at) = \exp(bt) \exp\left(\mu(at) + \frac{\sigma^2(at)^2}{2}\right) = \exp\left((a\mu + b)t + \frac{(a^2\sigma^2)t^2}{2}\right)$$

Since this is the moment generating function for a RV that is $\mathcal{N}(a\mu + b, a^2\sigma^2)$, we have shown that by the uniqueness of MGFs that $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

\square