Published in Towards Data Science

Aerin Kim

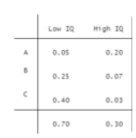
Sep 24, 2019 · 5 min read · Member-only · Listen

Moment Generating Function Explained

Its examples and properties

If you have Googled "Moment Generating Function" and the first, the second, and the third results haven't had you nodding yet, then give this article a try.

Aerin Kim
8.7K Followers
I'm an Engineering Manager at Scale AI and this is my notepad for Applied Math / CS / Deep Learning topics. Follow me on Twitter for more!



1. First things first — What is the "Moment" in probability/statistics?

Let's say the random variable we are interested in is **X**.

The moments are the expected values of X, e.g., E(X), E(X²), E(X³), … etc.

The first moment is **E(X)**,

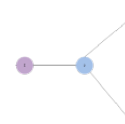
The second moment is **E(X²)**,

The third moment is **E(X³)**,

…

The n-th moment is **E(X^n)**.

We are pretty familiar with the first two moments, the mean **μ = E(X)** and the variance **E(X²) − μ²**. They are important characteristics of **X**.



The mean is the average value and the variance is how spread out the distribution is. But there must be **other features as well** that also define the distribution. For example, the third moment is about the asymmetry of a distribution. The fourth moment is about how heavy its tails are.

Moment	Uncentered	Centered
1st	$E(X) = \mu$	
2nd	$E(X^2)$	$E((X-\mu)^2)$
3rd	$E(X^3)$	$E((X-\mu)^3)$
4th	$E(X^4)$	$E((X-\mu)^4)$

$$\text{Mean}(X) = E(X)$$
$$\text{Var}(X) = E((X-\mu)^2) = \sigma^2$$
$$\text{Skewness}(X) = E((X-\mu)^3) / \sigma^3$$
$$\text{Kurtosis}(X) = E((X-\mu)^4) / \sigma^4$$

The **moments** tell you about the **features** of the distribution.

2. Then what is Moment Generating Function (MGF)?

As its name hints, MGF is literally the function that generates the moments — **E(X), E(X²), E(X³), … , E(X^n).**

$$MGF_X(t) := E[e^{tX}] = \begin{cases} \sum_x e^{tx} \cdot \overset{\text{PMF}}{p(x)} & X: \text{discrete} \\ \int_x e^{tx} \cdot \underset{\text{PDF}}{f(x)}\, dx & X: \text{continuous} \end{cases}$$

The definition of Moment-generating function

If you look at the definition of MGF, you might say…

> *"I'm not interested in knowing E(e^tx). I want E(X^n)."*

Take a derivative of MGF n times and plug t = 0 in. Then, you will get E(X^n).

$$\underset{\text{n-th moment}}{E(X^n)} = \frac{d^n}{dt^n} MGF_X(t) \Big|_{t=0}$$

e.g.

$$E(X) = \frac{d}{dt} MGF_X(t) \Big|_{t=0} = MGF'_X(0)$$

$$E(X^2) = \frac{d^2}{dt^2} MGF_X(t) \Big|_{t=0} = MGF''_X(0)$$

$$\vdots$$

This is **how you get the moments from the MGF.**

3. Show me the proof. Why is the n-th moment the n-th derivative of MGF?

We'll use Taylor series to prove this.

① $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$

then,

$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^n}{n!}$

Then take the expected value.

② $$E(e^{tx}) = E\left(1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^n}{n!}\right)$$
$$= E(1) + tE(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3}{3!}E(x^3) + \cdots + \frac{t^n}{n!}E(x^n)$$

Now, take a derivative with respect to **t**.

③ $$\frac{d}{dt}E(e^{tx}) = \frac{d}{dt}\left(E(1) + tE(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3}{3!}E(x^3) + \cdots + \frac{t^n}{n!}E(x^n)\right)$$

plug $t=0$ in...

$$= 0 + E(x) + 0 + 0 + \cdots + 0$$
$$= E(x) \quad \blacksquare$$

If you take another derivative on ③ (therefore total twice), you will get **E(X²)**.

If you take another (the third) derivative, you will get **E(X³)**, and so on and so on…

When I first saw the Moment

Generating Function, I couldn't understand **the role of t** in the function, because **t** seemed like some arbitrary variable that I'm not interested in. However, as you see, **t** is a helper variable. We introduced **t** in order to be able to use calculus (derivatives) and make the terms (that we are not interested in) zero.

> Wait… but we can calculate moments using the definition of expected values. Why do we need MGF exactly?

We can calculate $E(X^n)$ using the def. of Expected Value.

$$E(X^n) = \int_{-\infty}^{\infty} x^n \cdot \mathrm{pdf}(x)\, dx$$

The definition of Expected Value (Are you confused with **X** vs *x* notation? Check it out here.)

4. Why do we need MGF?

For convenience.

We want the MGF in order to calculate moments easily.

But why is the MGF easier than the definition of expected values?

In my math textbooks, they always told me to “**find the moment generating functions of Binomial(n, p), Poisson(λ), Exponential(λ), Normal(0, 1)**, etc.” However, they never really showed me why MGFs are going to be useful in such a way that they spark joy.

I think the below example will cause a spark of joy in you — **the clearest example where MGF is easier: The MGF of the exponential distribution**. (Don’t know what the exponential distribution is yet? 👉 The Intuition of Exponential Distribution)

We'll start with the PDF.

$$f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$

The PDF of exponential distribution

Deriving the MGF of exponential.

$$MGF_X(t) = E[e^{tx}] = \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\, dx$$

$\because x > 0$

$$= \lambda \underbrace{\int_0^\infty e^{(t-\lambda)x}\, dx}$$

for this to not blow up, what condition should be satisfied?

$t - \lambda < 0.$

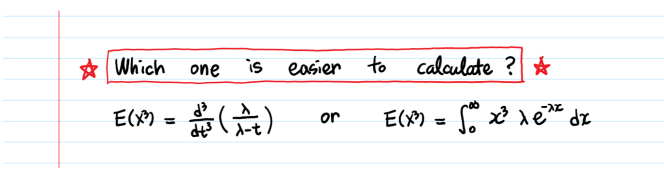
$$= \lambda \left| \frac{1}{t-\lambda} e^{(t-\lambda)x} \right|_0^\infty$$

$$= \lambda \left(0 - \frac{1}{t-\lambda}\right) = \frac{\lambda}{\lambda - t} \quad \blacksquare$$

SUPER SIMPLE !!!

For the MGF to exist, the expected value **E(e^tx)** should exist. This is why **`t - λ < 0`** is an important condition to meet, because otherwise the integral won't converge. (This is called **the divergence test** and is the first thing to check when trying to determine whether an integral converges or diverges.)

Once you have the MGF: **λ/(λ-t)**, **calculating moments becomes just a matter of taking derivatives**, which is easier than the integrals to calculate the expected value directly.



> Using MGF, it is possible to find moments by taking derivatives rather than doing integrals!

A few things to note:

1. For any valid MGF, $\mathbf{M(0) = 1}$. Whenever you compute an MGF, plug in $\mathbf{t = 0}$ and see if you get 1.
2. Moments provide a way to specify a distribution. For example, you can completely specify the normal distribution by the first two moments which are a mean and variance. As you know multiple different moments of the distribution, you will know more about that distribution. If there is a person that you haven't met, and you know about their height, weight, skin color, favorite hobby, etc., you still don't necessarily fully know them but are getting more and more information about them.
3. The beauty of MGF is, once you have MGF (once the expected value exists), you can get any n-th moment. MGF encodes all the

moments of a random variable into a single function from which they can be extracted again later.

4. A probability distribution is uniquely determined by its MGF. If two random variables have the same MGF, then they must have the same distribution. (Proof.)

5. For the people (like me) who are curious about the terminology "moments": Why is a moment called moment?

6. [Application 🔥] One of the important features of a distribution is how heavy its tails are, especially for risk management in finance. If you recall the 2009 financial crisis, that was essentially the failure to address the possibility of rare events happening. Risk managers understated the kurtosis (kurtosis means 'bulge' in Greek) of many financial securities underlying the fund's trading positions. Sometimes seemingly random distributions with hypothetically smooth curves of risk can have hidden

bulges in them. And we can detect those using MGF!