Moment Generating Functions

Eric Ingram

2024-11-16

erimingram.github.io

Mathematical Statistics: Moment Generating Functions (MGFs)

This document is a simple explanation of what MGFs are.

This document assumes that you have an understanding of basic mathematical statistics, derivatives/integrals, power series, what an open interval is, and mathematical notation. It will try to explain things in a simplified format: forgive the details left out.

This document does not cover higher-order moments (skewness, kurtosis, etc). If you are interested in that, I will make a separate document at some point.

This document also does not cover applications of MGFs and moments such as the method of moments, the MGF method for finding the distribution of a transformation, and the central limit theorem.

If you are a student in AI/CS/DSC 391L reading this, there is almost no relevance to the course other than that it can help you understand the theory behind some probability bounds such as the Chernoff bound.

What is a MGF?

A MGF is a function that generates all the moments of a random variable X. Specifically, the **kth moment** of X is defined as $E(X^k)$ for $k = 1, 2, \ldots$.

So, it should be pretty clear that the **first moment** is just the mean, $E(X^1) = E(X)$. The **second moment**, $E(X^2)$ can be used to calculate the variance in conjunction with the first moment.

The MGF of X, often denoted as $M_X(t)$ with dummy variable $t \in \mathbb{R}$, is defined as $M_X(t) = E(e^{tX})$.

In the discrete case, the MGF is calculated as $M_X(t) = E(e^{tX}) = \sum_x e^{tx} p(x)$, where the sum is over all possible values of X, and $p(x)$ is the pmf of X.

In the continuous case the MGF is calculated as $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)$ provided the integral converges. $f(X)$ is the pdf of X.

The MGF may not always exist. It exists if $E(e^{tX})$ is finite in some neighborhood (open interval) around $t = 0$. An important property of the MGF is that if it exists in an open interval containing $t = 0$, it uniquely determines the distribution of the random variable X.

We are particularly interested in the MGF in a neighborhood around $t = 0$ because the moments of X can be obtained by differentiating the MGF with respect to t and evaluating at $t = 0$: $E(X^k) = M_X^{(k)}(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$.

This allows us to generate all moments of X by successive differentiation.

Proof

Recall the power series of e^x: $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$.

Hence, $e^{tx} = \sum_{i=0}^{\infty} \frac{(tx)^i}{i!} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots$

So, if X is a discrete RV with pmf $p(x)$, $M_X(t) = E(e^{tX}) = \sum_x e^{tx} p(x) = \sum_x (1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots)p(x)$ (substituted in the power series).

Then, we can further simplify the above term: $M_X(t) = \sum_x (1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots)p(x) = \sum_x (p(x) + tx * p(x) + \frac{t^2x^2}{2!} p(x) + \frac{t^3x^3}{3!} p(x) + \ldots) = \sum_x p(x) + \sum_x tx * p(x) + \sum_x \frac{t^2x^2}{2!} p(x) + \sum_x \frac{t^3x^3}{3!} p(x) + \cdots = 1 + t \sum_x x * p(x) + \frac{t^2}{2!} \sum_x x^2 p(x) + \frac{t^3}{3!} \sum x^3 p(x) + \ldots$.

Recognize that these are moments!!! Therefore, $M_X(t) = 1 + t \sum_x x * p(x) + \frac{t^2}{2!} \sum_x x^2 p(x) + \frac{t^3}{3!} \sum x^3 p(x) + \cdots = 1 + t * E(X) + \frac{t^2}{2!} E(X^2) + \frac{t^3}{3!} E(X^3) + \ldots$.

Now, let's try differentiating this:

$\frac{d}{dt} M_X(t) = \frac{d}{dt}[1 + t * E(X) + \frac{t^2}{2!} E(X^2) + \frac{t^3}{3!} E(X^3) + \ldots] = E(X) + t * E(X^2) + \frac{t^2}{2!} E(X^3) + \ldots$.

Now, notice what happens if we plug in $t = 0$: $M_X'(0) = E(X) + 0 + 0 + \cdots = E(X)$. So, we generated the first moment by taking a first derivative and plugging in $t = 0$. Awesome!

This clearly extends to further derivatives and moments, giving us the result we claimed above: $E(X^k) = M_X^{(k)}(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$.

Basic Example:

Suppose $X \sim \text{Poisson}(\lambda)$. Find the MGF, mean, and variance of X.

MGF:

X is a Discrete RV here.

Therefore, the MGF of X is defined as $M_X(t) = E(e^{tX}) = \sum_x e^{tx} p(x)$.

Recall that the pmf of a Poisson RV is $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots$.

Let's substitute this into the MGF: $M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda}e^{\lambda e^t}$ since $\sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a$.

Then, $M_X(t) = e^{-\lambda}e^{\lambda e^t} = e^{-\lambda+\lambda e^t} = e^{\lambda(e^t-1)}$.

So, we have shown that the MGF of X is $e^{\lambda(e^t-1)}$.

Now that we have the MGF, let's find the mean and variance of X. We can always find the mean and variance the traditional way, but since we have the MGF we can do it with the MGF.

Mean:

Recall that $E(X) = M_X'(0)$. So, $M_X'(t) = \frac{d}{dt}[e^{\lambda(e^t-1)}] = e^{\lambda(e^t-1)}\lambda e^t = M_X(t)\lambda e^t$.

Plug in $t = 0$: $M_X'(0) = e^{\lambda(e^0-1)}\lambda e^0 = e^{\lambda(0)}\lambda = \lambda$.

Therefore, $E(X) = M_X'(0) = \lambda$.

Variance:

Now, recall that $Var(X) = E(X^2) - E(X)^2$. We have $E(X)$, the first moment. Now, we need to find $E(X^2)$, the second moment.

Realize that $E(X^2) = M_X''(0)$. In other words, the second derivative of the MGF, evaluated at $t = 0$. So, let's calculate it!

We already have the first derivative, so let's start there. $M_X''(t) = \frac{d}{dt}M_X'(t) = \frac{d}{dt}[M_X(t)\lambda e^t] = M_X'(t)\lambda e^t + M_X(t)\lambda e^t$ by the product rule.

Then, proceed by substituting in the known values: $M_X''(t) = M_X'(t)\lambda e^t + M_X(t)\lambda e^t = (M_X(t)\lambda e^t)\lambda e^t + M_X(t)\lambda e^t = M_X(t)\lambda e^t(\lambda e^t + 1)$.

Then, plug in $t = 0$: $M_X''(0) = M_X(0)\lambda e^0(\lambda e^0 + 1) = e^{\lambda(1-1)}\lambda * 1(\lambda * 1 + 1) = e^0 * \lambda(\lambda + 1) = \lambda(\lambda + 1)$.

Now, we can use the definition of variance: $Var(X) = E(X^2) - E(X)^2 = M_X''(0) - [M_X'(0)]^2 = \lambda(\lambda+1) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Hence, $Var(X) = \lambda$.

Properties

Translation Property

For a random variable X and a constant $b \in \mathbb{R}$, $M_{X+b}(t) = E(e^{t(X_b)}) = E(e^{tX+tb}) = E(e^{tX}e^{tb}) = e^{tb}E(e^{tX}) = e^{tb}M_X(t)$.

This says that when you shift a random variable X by a constant b to obtain $X + b$, the MGF of the new RV $X + b$ is related to the MGF of X by the factor e^{tb}.

Scaling Property

For a random variable X and a constant $b \in \mathbb{R} \setminus \{0\}$, $M_{aX}(t) = E(e^{t(aX)}) = E(e^{(at)X}) = M_X(at)$.

This says that when you scale a random variable X by a constant $a \neq 0$, the MGF of aX is the original MGF evaluated at at.

Additive Property

If $X_1, X_2, ..., X_n$ are independent random variables with MGFs $M_{X_i}(t)$ and $Y = \sum_{i=1}^{n} X_i$, then the MGF of Y is $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$.

This describes how the MGF of a sum of independent random variables relates to their individual MGFs. The MGF of the sum Y is the product of the MGFs of the individual X_is.

Uniqueness Property

As said before, MGFs are unique. So, if two random variables have the same MGF in an open interval containing $t = 0$, they have the same distribution.

This is what the MGF method for finding the probability distribution of a transformation is based off of. This will not be covered in this document, as it will be covered elsewhere.

MGFs of Common Distributions:

Discrete Distributions

- Discrete Uniform
  - $M_X(t) = \frac{1}{n}\sum_{k=a}^{b} e^{tk} = \frac{e^{t(b+1)} - e^{ta}}{n(e^t - 1)}$, over the integers $a, a+1, \ldots, b$.
- Bernoulli
  - $M_X(t) = E(e^{tX}) = (1-p)e^0 + pe^t = 1 - p + pe^t$
- Binomial
  - $M_X(t) = [(1-p) + pe^t]^n$
  - Derived from the additive property with n independent RVs.
- Geometric
  - $M_X(t) = \frac{p}{1-(1-p)e^t}, e^t < \frac{1}{1-p}$
  - The MGF only exists for t such that $e^t < \frac{1}{1-p}$
- Hypergeometric
  - The MGF does not have a simple closed-form expression due to the dependence between trials (since sampling is without replacement) which results in complex combinatorial sums. Calculate with software.
- Negative Binomial
  - $M_X(t) = (\frac{pe^t}{1-(1-p)e^t})^k, e^t < \frac{1}{1-p}$ for the definition where X is the total number of trials needed to achieve k successes.
  - $M_X(t) = (\frac{p}{1-(1-p)e^t})^k, e^t < \frac{1}{1-p}$ for the definition where X is the number of failures before achieving the kth success.
- Poisson
  - $M_X(t) = e^{\lambda(e^t - 1)}$

Continuous Distributions

- Continuous Uniform
  - $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}, t \neq 0$, over the interval $[a, b]$
  - At $t = 0$, $M_X(0) = 1$
- Normal/Gaussian
  - $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
  - The MGF exists for all real t since all moments of the Normal distribution exist.
- Multivariate Normal
  - $M_X(t) = e^{\mathbf{t}^T \mu + \frac{1}{2}\mathbf{t}^T \mathbf{\Sigma} \mathbf{t}}$, for vector $\mathbf{t} \in \mathbb{R}^k$
- Beta
  - The MGF of the beta distribution does not have a general closed-form expression.
  - The MGF only exists for specific cases (most notably when $\alpha, \beta \in \mathbb{Z}$) and generally doesn't simplify neatly.
- Exponential
  - $M_X(t) = \frac{\lambda}{\lambda - t}, t < \lambda$ for rate parameter λ
  - $M_X(t) = \frac{1}{1-\theta t}, t < \frac{1}{\theta}$ for scale parameter θ
- Gamma

- $M_X(t) = (\frac{\beta}{\beta - t})^{\alpha}, t < \beta$, for shape parameter α and rate parameter β
- $M_X(t) = (\frac{1}{1 - \theta t})^{k}, t < \frac{1}{\theta}$ for shape parameter k and scale parameter θ.