MTL108
Moment Generating Function (MGF)

Rahul Singh

Moment Generating Function (MGF)

The moment-generating function (MGF) is a powerful tool in probability theory that encapsulates all moments of a random variable in a single function. It is used to derive moments, identify distributions, and establish convergence properties.

Motivation: We know that Taylor series expansion of e^t is given by

$$e^t = \sum_{j=0}^{\infty} \frac{t^j}{j!} = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots.$$

Therefore,

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \mathbb{E}\left(1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \dots\right)\\ &= 1 + t\mathbb{E}(X) + \frac{t^2\mathbb{E}(X^2)}{2!} + \frac{t^3\mathbb{E}(X^3)}{3!} + \dots.\end{aligned}$$

Definition 1. For a random variable X, the Moment Generating Function (MGF), denoted $M_X(t)$, is defined as the expected value of e^{tX}, for all real numbers t for which the expectation exists.

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

- **For discrete random variables:** $M_X(t) = \sum_x e^{tx}\mathbb{P}(X = x)$
- **For continuous random variables:** $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$

The Moment Generating Function (MGF) is a general tool applicable to both discrete and continuous random variables.

Theorem 1. *If X and Y are independent random variables, the PGF of their sum, $Z = X + Y$, is the product of their individual PGFs, that is,*

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

Theorem 2 (Existence and Uniqueness (without proof))**.** *If $M_X(t)$ exists for t in some interval around 0, it uniquely determines the distribution of X.*

Theorem 3 (Moment Derivation). *The k-th moment of X is given by*

$$\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}.$$

Remark 1. If MGF exists for a random variable X, then $\mathbb{E}[X] = M_X'(0)$, $\mathbb{E}[X^2] = M_X''(0)$.

Theorem 4 (Linearity). *For constants a and b and random variables X and Y (if the MGFs exist),*

$$M_{aX+b}(t) = e^{bt} M_X(at) \quad \text{(for a single variable)},$$

and if X and Y are independent,

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

Theorem 5 (Convolution and Sums). *The MGF of the sum of independent random variables is the product of their MGFs, facilitating the analysis of combined distributions.*

Properties of the MGF

(i) **Generating Moments:** The k-th raw moment, $\mathbb{E}[X^k]$, is obtained by finding the k-th derivative of the MGF and evaluating it at $t = 0$.

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

For example, $\mathbb{E}[X] = M_X'(0)$ and $\mathbb{E}[X^2] = M_X''(0)$. The variance is then found using $\text{Var}(X) = M_X''(0) - (M_X'(0))^2$.

(ii) **Uniqueness Result:** If the MGF exists in an open interval around 0, it uniquely determines the probability distribution. This is a powerful property used to identify distributions.

(iii) **Linear Transformations:** For a linear transformation $Y = aX + b$, the MGF of Y is,

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

(iv) **Sum of Independent RVs:** If X and Y are independent, the MGF of their sum is the product of their individual MGFs.

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

Applications of MGFs

- **Proving the Central Limit Theorem:** The MGF of a standardized sum of independent, identically distributed random variables converges to the MGF of a standard normal distribution.

- **Convolution of Distributions:** Calculating the distribution of a sum of random variables is simplified by multiplying their MGFs instead of using convolution integrals.

Example 1 (Bernoulli distribution)**.** The PMF is $\mathbb{P}(X = 0) = 1 - p$, $\mathbb{P}(X = 1) = p$. The MGF is

$$M_X(t) = \sum_{x=0}^{1} e^{tx} p_X(x) = e^{t\cdot 0} \cdot (1-p) + e^{t\cdot 1} \cdot p = (1-p) + pe^t.$$

- Moments:

- First derivative: $M_X'(t) = pe^t$, so $\mathbb{E}[X] = M_X'(0) = p$,

- Second derivative: $M_X''(t) = pe^t$, so $\mathbb{E}[X^2] = M_X''(0) = p$,

- Variance: $\text{Var}(X) = p - p^2 = p(1-p)$.

This confirms the mean and variance derived earlier.

Example 2 (Exponential Distribution)**.** Let $X \sim \text{Exponential}(\lambda)$, with PDF $f(x) = \lambda e^{-\lambda x}$ for $x > 0$. The MGF is:

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{x(t-\lambda)}\, dx \\ &= \lambda \left[\frac{e^{x(t-\lambda)}}{t-\lambda} \right]_0^\infty \end{aligned}$$

This integral converges only if $t - \lambda < 0 \implies t < \lambda$.

$$\begin{aligned} M_X(t) &= \lambda \left(0 - \frac{e^0}{t-\lambda} \right) \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

Finding the Mean and Variance

- $M_X'(t) = \frac{d}{dt}\left(\lambda(\lambda-t)^{-1}\right) = \lambda(-1)(\lambda-t)^{-2}(-1) = \frac{\lambda}{(\lambda-t)^2}$ $\mathbb{E}[X] = M_X'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$
- $M_X''(t) = \frac{d}{dt}\left(\lambda(\lambda-t)^{-2}\right) = \lambda(-2)(\lambda-t)^{-3}(-1) = \frac{2\lambda}{(\lambda-t)^3}$ $\mathbb{E}[X^2] = M_X''(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$
- $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$

Table 1: Discrete distributions' summary

Distribution	Notation	PMF $P(X=k)$	Mean	Variance	MGF $M_X(t)$
Bernoulli	$X \sim \text{Bernoulli}(p)$	$\begin{cases} p, & k=1 \\ 1-p, & k=0 \end{cases}$	p	$p(1-p)$	$(1-p)+pe^t$
Binomial	$X \sim \text{Binomial}(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k},\ k=0,\ldots,n$	np	$np(1-p)$	$\left((1-p)+pe^t\right)^n$
Poisson	$X \sim \text{Poisson}(\lambda)$	$\frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,\ldots$	λ	λ	$\exp\left(\lambda(e^t-1)\right)$
Geometric	$X \sim \text{Geometric}(p)$	$(1-p)^{k-1}p,\ k=1,2,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t},\ t<-\ln(1-p)$
Negative Binomial	$X \sim \text{NegBin}(r,p)$	$\binom{k+r-1}{k}p^r(1-p)^k, \quad k=0,1,2,\ldots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r,\ t<-\ln(1-p).$
Uniform Discrete	$X \sim \text{UnifDis}(\{x_1,\ldots,x_m\})$	$1/m,\ k=x_1,\ldots,x_m$	$\frac{1}{m}\sum_{i=1}^m x_i$	$\frac{1}{m}\sum_{i=1}^m x_i^2 - \left(\frac{1}{m}\sum_{i=1}^m x_i\right)^2$	$\frac{1}{m}\sum_{i=1}^m e^{tx_i}$
Uniform Discrete	$X \sim \text{UnifDis}(\{1,\ldots,m\})$	$1/m,\ k=1,\ldots,m$	$(m+1)/2$	$(m^2-1)/12$	$\frac{1}{m}\sum_{i=1}^m e^{it}$

Table 2: Continuous distributions' summary

Distribution	Notation	PDF $f_X(x)$	Mean	Variance	MGF $M_X(t)$
Uniform	$X \sim \text{Uniform}(a,b)$	$\begin{cases} \frac{1}{b-a}, & \text{if} \quad a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$	$(b+a)/2$	$(b-a)^2/12$	$\frac{e^{tb}-e^{ta}}{t(b-a)},\ t \neq 0$
Exponential	$X \sim \text{Exponential}(\lambda)$	$\begin{cases} \lambda e^{-\lambda x}, & \text{if} \quad x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$	$1/\lambda$	$1/\lambda^2$	$\frac{\lambda}{\lambda-t},\ t<\lambda$
Normal	$X \sim \mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<\infty$	μ	σ^2	$e^{\mu t+\frac{\sigma^2 t^2}{2}}$
Standard Normal	$Z \sim \mathcal{N}(0,1)$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}},\ -\infty<z<\infty$	0	1	$e^{\frac{t^2}{2}}$

Theorem 6. *Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be two independent random variables. Then, their sum, $Z = X+Y$, is also normally distributed, specifically $Z \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$.*

Proof. Step 1: MGFs for X and Y
The MGF of a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is given by the formula:

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Using this formula for our independent random variables X and Y, MGFs are

$$M_X(t) = \exp\left(\mu_X t + \frac{1}{2}\sigma_X^2 t^2\right) \quad \text{and} \quad M_Y(t) = \exp\left(\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2\right)$$

Step 2: MGF of $Z = X+Y$
Since X and Y are independent, the MGF of their sum $Z = X+Y$ is the product of their individual MGFs.

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) \\ &= \exp\left(\mu_X t + \frac{1}{2}\sigma_X^2 t^2\right) \cdot \exp\left(\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2\right) \end{aligned}$$

Using the property of exponents, $e^a e^b = e^{a+b}$, we can combine the terms:

$$\begin{aligned} M_Z(t) &= \exp\left((\mu_X t + \frac{1}{2}\sigma_X^2 t^2) + (\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2)\right) \\ &= \exp\left((\mu_X + \mu_Y)t + \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2\right) \end{aligned}$$

Step 3: Distribution of Z

By comparing the derived MGF, $M_Z(t)$, with the general form of the MGF for a normal distribution from Step 1, we can see they are identical in form. The MGF $M_Z(t)$ matches the MGF of a normal distribution with:

$$\begin{aligned} \text{Mean} \quad & \mu_Z = \mu_X + \mu_Y \\ \text{Variance} \quad & \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \end{aligned}$$

By the uniqueness theorem for MGFs, if two distributions have the same MGF, they must be the same distribution. Therefore, $Z = X + Y$ is a normally distributed random variable.

$$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

□

Example 3. Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ be two independent random variables. Then, their sum, $Z = X + Y$, also follows a binomial distribution, specifically $Z \sim \text{Binomial}(n + m, p)$.

Proof using Moment Generating Function (MGF)

1. MGF of a Binomial RV: The MGF of a random variable X is defined as $M_X(t) = E(e^{tX})$. For a binomial random variable $X \sim \text{Binomial}(n, p)$, the PMF is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$. Thus, the MGF for X is:

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \sum_{k=0}^{n} e^{tk} P(X = k) \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \end{aligned}$$

By the binomial theorem, $\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n$. Letting $a = pe^t$ and $b = 1 - p$, we get:

$$M_X(t) = ((1-p) + pe^t)^n$$

2. MGF of $X + Y$: Given that X and Y are independent, the MGF of their sum $Z = X + Y$ is the product of their individual MGFs.

$$
\begin{aligned}
M_Z(t) &= M_X(t)M_Y(t) \\
&= ((1-p)+pe^t)^n \cdot ((1-p)+pe^t)^m \\
&= ((1-p)+pe^t)^{n+m}
\end{aligned}
$$

3. Conclusion: The resulting MGF, $M_Z(t) = ((1-p)+pe^t)^{n+m}$, is the MGF of a binomial distribution with parameters $n+m$ and p. Since the MGF uniquely identifies a probability distribution, we can conclude that $Z \sim \text{Binomial}(n+m, p)$.

Optional

Mixture Distribution

A mixture random variable is one whose distribution is a mixture of other distributions. This often arises when a random variable is chosen from a set of different types, each with its own probability.

Example 4. Consider a factory where items are produced by two different machines, Machine A and Machine B. Machine A produces items with a defect rate of p_A, and Machine B produces items with a defect rate of p_B. The factory uses Machine A with probability q and Machine B with probability $1-q$. Let X be the number of defects in an item produced by a randomly chosen machine. X can be modeled as a mixture of two Bernoulli distributions.

$$
X = \begin{cases} X_A \sim \text{Bernoulli}(p_A) & \text{with probability } q \\ X_B \sim \text{Bernoulli}(p_B) & \text{with probability } 1-q \end{cases}
$$

Finding the MGF of the Mixture Distribution

The MGF of the mixture variable X is the weighted average of the MGFs of the component distributions.

$$M_X(t) = q \cdot M_{X_A}(t) + (1-q) \cdot M_{X_B}(t)$$

The MGF for a Bernoulli variable $B \sim \text{Bernoulli}(p)$ is $M_B(t) = (1-p)e^{t\cdot 0} + pe^{t\cdot 1} = 1-p+pe^t$. Therefore, the MGF for our mixture distribution is:

$$M_X(t) = q(1-p_A+p_Ae^t) + (1-q)(1-p_B+p_Be^t)$$

This is a key application of MGFs in understanding the overall behavior of a system with multiple component behaviors.

Finding the Mean of the Mixture Distribution

We can find the mean of the mixture distribution by differentiating its MGF and evaluating it at $t = 0$.

$$\begin{aligned} M'_X(t) &= q(p_A e^t) + (1-q)(p_B e^t) \\ \mathbb{E}[X] &= M'_X(0) = q(p_A e^0) + (1-q)(p_B e^0) \\ &= qp_A + (1-q)p_B \end{aligned}$$

This result makes intuitive sense: the overall defect rate is the weighted average of the defect rates of the two machines.

Example 5 (Neither discrete nor continuous)**.** Let X be a random variable representing the payout amount for an insurance policy. The cumulative distribution function (CDF) of X, denoted by $F_X(x)$, is a mixed distribution consisting of a discrete part and a continuous part.

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ w_0 & \text{if } x = 0 \\ w_0 + (1-w_0)F_C(x) & \text{if } x > 0. \end{cases}$$

Here w_0 is the probability of a zero payout (no claim), and $F_C(x)$ is the cumulative distribution function of the continuous random variable for positive payouts. For example, if we model the positive claims with a continuous random variable C that follows a log-normal distribution, the continuous part of the CDF would be $F_C(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$ for $x > 0$.

Optional

Probability Generating Function (PGF)

Definition 2. The Probability Generating Function (PGF) is a tool for analyzing discrete random variables taking non-negative integer values. For such a random variable X, the PGF, denoted by $G_X(t)$, is a function of a real t defined as the expected value of t^X.

$$G_X(t) = \mathbb{E}[t^X] = \sum_{x=0}^{\infty} t^x \mathbb{P}(X = x)$$

This power series representation conveniently encodes all the probabilities of the distribution, with the coefficient of t^x being the probability $\mathbb{P}(X = x)$.

Remark 2. The PGF is simply a probability distribution compressed into a function. The coefficients of the polynomial (or power series) expansion of $G_X(t)$ give the probabilities.

Theorem 7. *If X and Y are independent random variables, the PGF of their sum, $Z = X + Y$, is the product of their individual PGFs, that is,*

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t).$$

Proof. Observe that

$$G_{X+Y}(t) = \mathbb{E}(t^{X+Y}) = \mathbb{E}(t^X \cdot t^Y).$$

Since t^X is a function of X only and t^Y is a function of Y only, using independence and Theorem 3 in Topic 8, we have

$$G_{X+Y}(t) = \mathbb{E}(t^X \cdot t^Y) = \mathbb{E}(t^X) \cdot \mathbb{E}(t^Y) = G_X(t) \cdot G_Y(t).$$

□

Properties of the PGF

(i) **Normalisation:** Substituting $t = 1$ into the PGF gives the sum of all probabilities, which must equal 1.

$$G_X(1) = \sum_{x=0}^{\infty} (1)^x \mathbb{P}(X = x) = \sum_{x=0}^{\infty} \mathbb{P}(X = x) = 1$$

This can be used to find unknown constants in a PGF.

(ii) **Generating Probabilities:** The probability mass function can be recovered from the PGF using its derivatives.

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}$$

For example, $\mathbb{P}(X = 0) = G_X(0)$ and $\mathbb{P}(X = 1) = G'_X(0)$.

(iii) **Finding Moments:** The PGF can be used to find the factorial moments of X.

- The first derivative evaluated at $t = 1$ gives the mean.

$$\mathbb{E}[X] = G'_X(1)$$

- The second derivative evaluated at $t = 1$ gives the second factorial moment, $\mathbb{E}[X(X-1)]$.

$$\mathbb{E}[X(X-1)] = G''_X(1)$$

The variance can then be found using the identity:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = G''_X(1) + G'_X(1) - (G'_X(1))^2$$

(iv) **Uniqueness:** The PGF uniquely determines the probability distribution. If two random variables have the same PGF, they must have the same distribution.

(v) **Sum of Independent RVs:** If X and Y are independent random variables, the PGF of their sum, $Z = X + Y$, is the product of their individual PGFs.

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t)$$

Example 6 (Poisson Distribution). Let $X \sim \text{Poisson}(\lambda)$, with probability mass function $\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. The PGF is:

$$\begin{aligned} G_X(t) = \mathbb{E}[t^X] &= \sum_{k=0}^{\infty} t^k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda t} \\ &= e^{\lambda(t-1)} \end{aligned}$$

Finding the probability from the PGF

$$\begin{aligned} G'_X(t) &= \frac{d}{dt} e^{\lambda(t-1)} = \lambda e^{\lambda(t-1)} \\ \Rightarrow \mathbb{P}(X = 1) &= G'_X(0) = \lambda e^{\lambda(0-1)} = \lambda e^{-\lambda} \\ \text{Next} \quad G''_X(t) &= \frac{d}{dt} G'_X(t) = \frac{d}{dt}[\lambda e^{\lambda(t-1)}] = \lambda^2 e^{\lambda(t-1)} \\ \mathbb{P}(X = 2) &= G''_X(0)/2! = \lambda^2 e^{\lambda(0-1)}/2! = e^{-\lambda}\lambda^2/2!. \end{aligned}$$

Finding the Mean from the PGF

$$\begin{aligned} G'_X(t) &= \frac{d}{dt} e^{\lambda(t-1)} = \lambda e^{\lambda(t-1)} \\ \mathbb{E}[X] &= G'_X(1) = \lambda e^{\lambda(1-1)} = \lambda e^0 = \lambda \end{aligned}$$

This confirms that the mean of a Poisson(λ) distribution is λ.

Relationship between PGF and MGF

For a discrete, non-negative integer-valued random variable X, the MGF and PGF are directly related by the substitution $t \to e^t$ in the PGF.

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \mathbb{E}[(e^t)^X] \\ &= G_X(e^t) \end{aligned}$$

Conversely, using $t^X = e^{(\log t)X}$, we have

$$\begin{aligned} G_X(t) &= \mathbb{E}[t^X] = \mathbb{E}[e^{(\log t)X}] \\ &= M_X(\log t). \end{aligned}$$

Applications of PGFs

- **Queueing Theory and Stochastic Processes:** PGFs are used to analyze the number of arrivals or customers in a system over time.

- **Branching Processes:** They are fundamental for studying population growth models where individuals reproduce independently.

Comparison Summary

Feature	**Probability Generating Function (PGF)**	**Moment Generating Function (MGF)**
Applicable to	Discrete RVs on non-negative integers only.	Both discrete and continuous RVs.
Definition	$G_X(t) = \mathbb{E}[t^X]$	$M_X(t) = \mathbb{E}[e^{tX}]$
Moment Generation	Factorial moments via derivatives at $t = 1$.	Ordinary (raw) moments via derivatives at $t = 0$.
Uniqueness	PGF uniquely determines distribution.	MGF uniquely determines distribution (if it exists).
Existence	Always exists for $t \in [-1, 1]$.	May not exist for all t.
Sum of Independent RVs	$G_{X+Y}(t) = G_X(t) \cdot G_Y(t)$	$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$
Relationship	For integer-valued X, $G_X(t) = M_X(\ln t)$.	$M_X(t) = G_X(e^t)$.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer