# Lecture 5: Moment generating functions

## Definition 2.3.6.

The moment generating function (mgf) of a random variable $X$ is

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f_X(x) & \text{if } X \text{ has a pmf} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \text{if } X \text{ has a pdf} \end{cases}$$

provided that $E(e^{tX})$ exists. (Note that $M_X(0) = E(e^{0X}) = 1$ always exists.) Otherwise, we say that the mgf $M_X(t)$ does not exist at $t$.

## Theorem 2.3.15.

For any constants $a$ and $b$, the mgf of the random variable $aX + b$ is

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

## Proof.

By definition,

$$M_{aX+b}(t) = E(e^{t(aX+b)}) = E(e^{taX} e^{bt}) = e^{bt} E(e^{(ta)X}) = e^{bt} M_X(at)$$

## The main use of mgf

- It can be used to generate moments.
- It helps to characterize a distribution.

## Theorem 2.3.7.

If $M_X(t)$ exists at $\pm t$, then $E(X^n)$ exists for any positive integer $n$ and

$$E(X^n) = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}$$

i.e., the $n$th moment is the $n$th derivative of $M_X(t)$ evaluated at $t = 0$.

## Proof.

Assuming that we can exchange the differentiation and integration (which will be justified later),

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

Hence

$$\frac{d^n}{dt^n} M_X(t) \Big|_{t=0} = \int_{-\infty}^{\infty} x^n e^{0x} f_X(x) dx = \int_{-\infty}^{\infty} x^n f_X(x) dx = E(X^n)$$

The condition that $M_X(t)$ exists at $\pm t$ is in fact used to ensure the validity of the exchange of differentiation and integration.

Besides this, we also need to argue that $E|X|^n < \infty$ for any positive integer $n$ under the condition that $M_X(t)$ exists at $\pm t$.

We first show that, if $M_X(t)$ exists at $\pm t$, then for any $s \in (-t,t)$, $M_X(s)$ exists:

$$\begin{aligned} M_X(s) &= E(e^{sX}) = E[e^{sX} I(X > 0)] + E[e^{sX} I(X \le 0)] \\ &\le E[e^{tX} I(X > 0)] + E[e^{-tX} I(X \le 0)] \le E(e^{tX}) + E(e^{-tX}) \\ &= M_X(t) + M_X(-t) < \infty \end{aligned}$$

where $I(X > 0)$ is the indicator of $X > 0$.

Next, we show that, for any positive $p > 0$, $E|X|^p < \infty$ under the condition $M_X(t)$ exists at $\pm t$.

For a given $p > 0$, choose $s$ such that $0 < ps < t$.

Because $s|X| \le e^{s|X|}$, we have $s^p|X|^p \le e^{ps|X|} \le e^{psX} + e^{-psX}$ and

$$E|X|^p \le s^{-p} E(e^{psX} + e^{-psX}) = s^{-p} M_X(ps) + M_X(-ps) < \infty$$

In fact, the condition that $M_X(t)$ exists at $\pm t$ ensures that $M_X(s)$ has the power series expansion

$$M_X(s) = \sum_{k=0}^{\infty} \frac{E(X^k)s^k}{k!} \qquad -t < s < t$$

If the distribution of $X$ is symmetric (about 0), i.e., $X$ and $-X$ have the same distribution, then

$$M_X(t) = E(e^{tX}) = E(e^{t(-X)}) = E(e^{-tX}) = M_X(-t)$$

i.e., $M_X(t)$ is an even function and $M_X(t)$ exists at $\pm t$ is the same as $M_X(t)$ exists at a $t > 0$.

## Example 2.3.8 (Gamma mgf)

Let $X$ have the gamma pdf

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

where $\alpha > 0$ and $\beta > 0$ are two constants and

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \qquad \alpha > 0$$

is the so-called gamma function.

If $t < 1/\beta$,

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{tx}x^{\alpha-1}e^{-x/\beta}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-x/(\frac{\beta}{1-\beta t})}dx \\ &= \frac{(\frac{\beta}{1-\beta t})^\alpha}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty s^{\alpha-1}e^{-s}ds = \frac{1}{(1-\beta t)^\alpha} \end{aligned}$$

If $t \geq 1/\beta$, then $E(e^{tX}) = \infty$.

We can obtain

$$E(X) = \frac{d}{dt}M_X(t)\bigg|_{t=0} = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}\bigg|_{t=0} = \alpha\beta$$

For any integer $n > 1$,

$$\begin{aligned} E(X^n) &= \frac{d^n}{dt^n}M_X(t)\bigg|_{t=0} = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta^n}{(1-\beta t)^{\alpha+n}}\bigg|_{t=0} \\ &= \alpha(\alpha+1)\cdots(\alpha+n-1)\beta^n \end{aligned}$$

## Can the moments determine a distribution?

Can two random variables with different distributions have the same moments of any order?

## Example 2.3.10.

$$X_1 \text{ has pdf} \quad f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, \quad x \geq 0$$

$$X_2 \text{ has pdf} \quad f_2(x) = f_1(x)[1 + \sin(2\pi \log x)], \quad x \geq 0$$

For any positive integer $n$,

$$\begin{aligned} E(X_1^n) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny - y^2/2} dy \qquad y = \log x \\ &= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(y-n)^2/2} dy \\ &= e^{n^2/2} \qquad \text{using the property of a normal distribution} \end{aligned}$$

$$
\begin{aligned}
E(X_2^n) &= \int_0^\infty x^n f_1(x)[1+\sin(2\pi \log x)]dx \\
&= E(X_1^n) + \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} \sin(2\pi \log x) dx \\
&= E(X_1^n) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny} e^{-y^2/2} \sin(2\pi y) dy \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(y-n)^2/2} \sin(2\pi y) dy \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi(s+n)) ds \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi s) ds \\
&= E(X_1^n)
\end{aligned}
$$

since $e^{-s^2/2}\sin(2\pi s)$ is an odd function.
This shows that $X_1$ and $X_2$ have the same moments of order $n = 1, 2, ...$, but they have different distributions.

In some cases, moments determine the distributions.
The mgf, if it exists, determines a distribution.

## Theorem 2.3.11

Let $X$ and $Y$ be random variables with cdfs $F_X$ and $F_Y$, respectively.

a. If $X$ and $Y$ are bounded, then $F_X(u) = F_Y(u)$ for all $u$ iff $E(X^r) = E(Y^r)$ for all $r = 1, 2, ...$
b. If mgf's exist in a neighborhood of 0 and $M_X(t) = M_Y(t)$ for all $t$, then $F_X(u) = F_Y(u)$ for all $u$.

The key idea of the proof can be explained as follows.

Note that

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

is the Laplace transformation of $f_X(x)$.

From the uniqueness of the Laplace transformation, there is a one-to-one correspondence between the mgf and the pdf.

We will give a proof of this result in Chapter 4 for the multivariate case, after we introduce the characteristic functions.

From the power series result in the last lecture, if the mgf of $X$ exists in a neighborhood of 0, then it has a power series expansion which is determined by moments $E(X^n)$, $n = 1, 2, \ldots$.

Therefore, knowing the mgf and knowing moments of all order are the same, but this is under the condition that the mgf exists in a neighborhood of 0.

Once we establish part (b), the proof of part (a) is easy: if $X$ and $Y$ are bounded, then their mgf's exist for all $t$ and thus their cdf's are the same iff their moments are the same for any order.

The condition that the mgf exists in a neighborhood of 0 is important.

There are random variables with finite moments of any order, but their mgf's do not exist.

## Example

The pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2}, \quad x \geq 0$$

is called the log-normal distribution or density, because if $X$ has pdf $f_X$, then $\log X$ has a normal pdf.

In Example 2.3.10, we have shown that the log-normal distribution has finite moments of any order.

For $t > 0$,

$$M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx = \infty$$

because, when $t > 0$,

$$\lim_{x\to\infty} \frac{e^{tx}}{\sqrt{2\pi}x} e^{-(\log x)^2/2} = \infty$$

When $t < 0$,

$$M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx \le \int_0^\infty \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx = 1$$

and, hence, $M_X(t)$ exists for all $t < 0$.

How do we find a distribution for which all moments exist and the mgf does not exists for any $t \neq 0$?

Consider the pdf

$$f_Y(x) = \begin{cases} f_X(x)/2 & x > 0 \\ f_X(-x)/2 & x < 0 \end{cases}$$

For this pdf,

$$E(|Y|^n) = \int_0^\infty x^n \frac{f_X(x)}{2} dx + \int_{-\infty}^0 (-x)^n \frac{f_X(-x)}{2} dx = \int_0^\infty x^n f_X(x) dx = E(X^n)$$

which has been derived for any $n = 1, 2, \ldots$

On the other hand,

$$\begin{aligned} E(e^{tY}) &= \int_0^\infty e^{tx} \frac{f_X(x)}{2} dx + \int_{-\infty}^0 e^{tx} \frac{f_X(-x)}{2} dx \\ &= \int_0^\infty e^{tx} \frac{f_X(x)}{2} dx + \int_0^\infty e^{-tx} \frac{f_X(x)}{2} dx \end{aligned}$$

and we have shown that one of these integrals is $\infty$ (depending on whether $t > 0$ or $< 0$).

## Theorem.

If a random variable $X$ has finite moment $a_n = E(X^n)$ for any $n = 1, 2, \ldots$, and the series

$$\sum_{n=0}^\infty \frac{|a_n||t|^n}{n!} < \infty \qquad \text{with } |t| > 0$$

then the cdf of $X$ is determined by $a_n$, $n = 1, 2 \ldots$

## Example.

Suppose that $a_n = n!$ is the $n$th moment of a random variable $X$.

Since
$$\sum_{n=0}^{\infty} \frac{|a_n||t|^n}{n!} = \sum_{n=0}^{\infty} |t|^n = \frac{1}{1-|t|} \qquad |t| < 1$$
and this function is the mgf of $Gamma(1,1)$ at $|t|$, we conclude that $X \sim Gamma(1,1)$.

Suppose that the $n$th moment of a random variable $Y$ is
$$a_n = \begin{cases} n!/(n/2)! & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

Then
$$\sum_{n=0}^{\infty} \frac{|a_n||t|^n}{n!} = \sum_{n \text{ is even}} \frac{n!(t^2)^{n/2}}{n!(n/2)!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{k!} = e^{t^2} \quad t \in \mathscr{R}$$
Later, we show that this is the mgf of $N(0,\sqrt{2})$, hence $X \sim N(0,\sqrt{2})$.

For a log-normal distributed random variable $X$ discussed in the beginning of this lecture, $E(X^n) = e^{n^2/2}$ and $\sum_{n=0}^{\infty} e^{n^2/2}|t|^n/n! = \infty$ for any $|t| > 0$ and, hence, the theorem is not applicable.

In applications we often need to approximate a cdf by a sequence of cdf's.

The next theorem gives a sufficient condition for the convergence of cdf's and moments of random variables in terms of the convergence of mgf's.

## Theorem 2.3.12

Suppose that $X_1, X_2, \ldots$ is a sequence of random variables with mgf's $M_{X_n}(t)$, and

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) < \infty \quad \text{for all } t \text{ in a neighborhood of } 0$$

where $M_X(t)$ is the mgf of a random variable $X$.
Then, for all $x$ at which $F_X(x)$ is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Furthermore, for any $p > 0$, we have

$$\lim_{n\to\infty} E|X_n|^p = E|X|^p \quad \text{and} \quad \lim_{n\to\infty} E|X_n - X|^p = 0$$

## Example 2.3.13 (Poisson approximation)

The cdf of the binomial pmf,

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n,$$

where $n$ is a positive integer and $0 < p < 1$, may not be easy to calculate when $n$ is very large.

It is often approximated by the cdf of the Poisson pmf,

$$f_Y(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

where $\lambda > 0$ is a constant.

For this purpose, we first compute the mgf's for the binomial and Poisson distributions.

## Example 2.3.9 (binomial mgf)

Using the binomial formula

$$\sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x} = (u+v)^n$$

we obtain

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

Note that

$$
\begin{aligned}
E(X) &= \frac{d}{dt} M_X(t) \Big|_{t=0} = n(pe^t + 1 - p)^{n-1} pe^t \Big|_{t=0} = np \\
E(X^2) &= \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} \\
&= [n(n-1)(pe^t + 1 - p)^{n-1} (pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t] \Big|_{t=0} \\
&= n(n-1)p^2 + np
\end{aligned}
$$

We got the same results previously, but the calculation here is simpler.

## Example 2.3.13 (continued)

If $Y$ has the Poisson pmf, then

$$M_Y(t) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t-1)}$$

which is finite for any $t \in \mathscr{R}$.

Let $X_n$ have the binomial distribution with $n$ and $p$.

Suppose that $\lim_{n\to\infty} np = \lambda > 0$ (that means $p$ also depends on $n$ and $p \to 0$ when $n \to \infty$).

Then, for any $t$, as $n \to \infty$,

$$M_{X_n}(t) = (pe^t + 1 - p)^n = \left[1 + \frac{(np)(e^t - 1)}{n}\right]^n \to e^{\lambda(e^t-1)} = M_Y(t)$$

using the fact that, for any sequence of numbers $a_n$ converges to $a$,

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

With this result and Theorem 2.3.12, we can approximate $P(X_n \leq u)$ by $P(Y \leq u)$ when $n$ is large and $np$ is approximately a constant.