Moment Generating Functions

Definition: The moment generating function (MGF) of X, written as $\psi(t)$, is

$$\psi(t) = E[e^{tX}].$$

When the expectation is not well-defined for some $t = t_0$, then we do not talk about $\psi(t_0)$.

If X is a continuous random variable with pdf $f(x)$, then

$$\psi(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

If X is a discrete random variable with probability function f, then

$$\psi(t) = \sum_x e^{tx} f(x).$$

Remark: When $t = 0$, MGF is always well defined and equals 1. $\psi(0) = 1$, because

$$\psi(0) = E[e^{0X}] = E[1] = 1.$$

Ex1

Let $X \sim Ber(p)$, which means X is either 1 or 0, and $P(X = 1) = p$, $P(X = 0) = 1 - p$. The moment generating function (MGF) of X is

$$\begin{aligned} \psi(t) &= E[e^{tX}] \\ &= e^{t \cdot 1} P(X = 1) + e^{t \cdot 0} P(X = 0) \\ &= e^t p + e^0 (1 - p) \\ &= 1 - p + pe^t \end{aligned}$$

The above equation holds for all $-\infty < t < \infty$.

Ex2

Let $X \sim Bin(n, p)$, which means X takes values from $0, 1, 2, \ldots, n$, and $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$. The moment generating function (MGF) of X is

$$
\begin{aligned}
\psi(t) &= E[e^{tX}] \\
&= \sum_{k=0}^{n} e^{t \cdot k} P(X = k) \\
&= \sum_{k=0}^{n} e^{t \cdot k} \binom{n}{k} p^k (1 - p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (e^t)^k p^k (1 - p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1 - p)^{n-k} \\
&= (e^t p + 1 - p)^n
\end{aligned}
$$

where in the last step we applied the Binomial theorem.

The above equation holds for all $-\infty < t < \infty$.

Ex3

Let X has the pdf $f(x) = e^{-x} I(x > 0)$.

$$\begin{aligned}
\psi(t) &= E[e^{tX}] \\
&= \int_0^\infty e^{tx} e^{-x} dx \\
&= \int_0^\infty e^{(t-1)x} dx
\end{aligned}$$

If $t \geq 1$, then $e^{(t-1)x} \geq 1$ for all $x \geq 0$. Thus

$$\int_0^\infty e^{(t-1)x} dx = \infty.$$

If $t < 1$,

$$\begin{aligned}
\psi(t) &= \int_0^\infty e^{(t-1)x} dx \\
&= \frac{1}{t-1} e^{(t-1)x} |_0^\infty \\
&= 0 - \frac{1}{t-1} \\
&= \frac{1}{1-t}
\end{aligned}$$

Ex4

Let $X \sim N(0,1)$. Find $\psi(t)$.

$$\begin{aligned}
\psi(t) &= E[e^{tX}] \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2-2tx}{2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2-t^2}{2}} dx \\
&= e^{\frac{1}{2}t^2} [\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx] \\
&= e^{\frac{1}{2}t^2} [\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds] \quad [s = x - t] \\
&= e^{\frac{1}{2}t^2} \cdot 1 \\
&= e^{\frac{1}{2}t^2}
\end{aligned}$$

Compute Moments by MGF

The k-th moment of X is defined as $E[X^k]$ for $k = 1, 2, 3, 4, \ldots$ whenever the expectation is well defined.

When the MGF of X exists for all t within a small open interval around 0 ($(-\delta, \delta)$, for some $\delta > 0$), then $E[X^k]$ exists for all $k \geq 1$ and

$$E[X^k] = \psi^{(k)}(0).$$

$\psi^{(k)}(0)$ is the k-th derivative of $\psi(t)$ at $t = 0$.

Explanation: By the Taylor's expansion of the exponential function we have

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

$$\begin{aligned}
\psi(t) &= E[e^{tX}] \\
&= E[1 + tX + \frac{1}{2!}(tX)^2 + \frac{1}{3!}(tX)^3 + \cdots] \\
&= 1 + tE[X] + \frac{1}{2!}t^2E[X^2] + \frac{1}{3!}t^3E[X^3] + \cdots
\end{aligned}$$

The Taylor's expansion of $\psi(t)$ gives us

$$\psi(t) = \psi(0) + t\psi^{(1)}(0)t + \frac{1}{2!}t^2\psi^{(2)}(0) + \frac{1}{3!}t^3\psi^{(3)}(0) + \cdots$$

Compare the two series we see

$$\psi^{(k)}(0) = E[X^k], \quad k = 1, 2, 3, \ldots.$$

Two special cases are of particular interest

$$\psi'(0) = E[X]; \; \psi''(0) = E[X^2].$$

where $\psi'(0) = \psi^{(1)}(0)$ and $\psi''(0) = \psi^{(2)}(0)$.

Therefore

$$E[X] = \psi'(0).$$

$$Var(X) = \psi''(0) - [\psi'(0)]^2.$$

Ex1

Let $X \sim Bin(n, p)$. Compute EX and $Var(X)$.

We have computed the MGF of X

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

$$\psi'(t) = n(1 - p + pe^t)^{n-1} pe^t$$

$$\begin{aligned} \psi''(t) &= \frac{d\left(n(1 - p + pe^t)^{n-1} pe^t\right)}{dt} \\ &= n(n-1)(1 - p + pe^t)^{n-2} pe^t pe^t \\ &\quad + n(1 - p + pe^t)^{n-1} pe^t \end{aligned}$$

$$EX = \psi'(0) = n(1 - p + pe^0)^{n-1} pe^0 = np.$$

$$\begin{aligned} \psi''(0) &= n(n-1)(1 - p + pe^0)^{n-2} pe^0 pe^0 \\ &\quad + n(1 - p + pe^0)^{n-1} pe^0 \\ &= n(n-1)p^2 + np \end{aligned}$$

$$\begin{aligned} Var(X) &= \psi''(0) - [\psi'(0)]^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= -np^2 + np = np(1 - p). \end{aligned}$$

Ex2

Let X has the pdf $f(x) = e^{-x} I(x > 0)$.

We have computed

$$\psi(t) \;=\; \frac{1}{1-t} \quad \text{for } t < 1.$$

$$\psi'(t) \;= \tfrac{1}{(1-t)^2}$$

$$\psi''(t) \;= \tfrac{2}{(1-t)^3}$$

$$\psi'''(t) \;= \tfrac{6}{(1-t)^4}$$

$$EX = \psi'(0) \;= 1$$

$$E[X^2] = \psi''(0) \;= 2$$

$$E[X^3] = \psi'''(0) \;= 6$$

$$Var(X) = E[X^2] - (EX)^2 = 2 - 1^2 = 1.$$

A useful theorem

Theorem: Suppose $X_1, X_2, \ldots, X_n$ are independent random variables and $\psi_{X_i}(t)$ is the same MGF of X_i. Let

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b.$$

Then the MGF of Y is

$$\psi_Y(t) = e^{tb} \prod_{i=1}^{n} \psi_{X_i}(a_i t).$$

Proof:

$$\begin{aligned}
\psi_Y(t) &= E[e^{tY}] \\
&= E[e^{ta_1X_1 + ta_2X_2 + \cdots + ta_nX_n + b}] \\
&= e^{tb} E[e^{ta_1X_1} e^{ta_2X_2} \cdots e^{ta_nX_n}] \\
&\quad \text{by independence assumption} \\
&= e^{tb} E[e^{ta_1X_1}] E[e^{ta_2X_2}] \cdots E[e^{ta_nX_n}] \\
&= e^{tb} \psi_{X_1}(ta_1) \psi_{X_2}(ta_2) \cdots \psi_{X_i}(ta_n)
\end{aligned}$$

Ex1

If X_is $1 \le i \le n$ are iid random variables and $Y = \frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}$. Find $\psi_Y(t)$.

Solution: We use the theorem. Note that $a_1 = \cdots = a_n = 1/\sqrt{n}$ and $b = 0$ so

$$\psi_Y(t) = \prod_{i=1}^{n} \psi_{X_i}(t/\sqrt{n}).$$

Furthermore, because X_is have the same distribution, $\psi_{X_1} = \psi_{X_2} = \cdots = \psi_{X_n}$, then

$$\psi_Y(t) = \psi_{X_1}(t/\sqrt{n})^n.$$

Ex2

Let $Y = \sigma X + \mu$ with $\sigma > 0$. Find $\psi_Y(t)$. Write down the explicit expression if assume X is a standard normal RV.

Solution: By the theorem we know

$$\psi_Y(t) = e^{t\mu}\psi_X(\sigma t).$$

If $X \sim N(0,1)$ then $\psi_X(t) = e^{\frac{1}{2}t^2}$. Hence

$$\psi_Y(t) = e^{t\mu}e^{\frac{1}{2}(\sigma t)^2} = e^{t\mu+\frac{\sigma^2}{2}t^2}.$$

Use MGFs to determine the distribution

First, if two random variables X and Y have the same distribution, that is, they have the same probability function or the same probability density function, then X and Y have the same MGF. This can be seen from the definition of MGF.

On the other hand, if X and Y have the same MGF, then they must follow the same distribution.

We will use this property to figure out the distribution of some random variable by comparing its MGF with the MGFs of existing known distributions.

Ex1

Let $X_1, X_2, \ldots, X_n$ are iid $Ber(p)$. Let $Y = X_1 + X_2 + \cdots + X_n$. Find the distribution of Y.

The MGF of X_i is $1 - p + pe^t$. $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ with $a_i = 1$

$$\psi_Y(t) = \prod_{i=1}^{n} \psi(t) = \psi(t)^n = (1 - p + pe^t)^n$$

Note that $(1-p+pe^t)^n$ is the MGF of $Bin(n,p)$. Thus we know $Y \sim Bin(n,p)$.

From this example we see that $Bin(n,p)$ can be viewed as sum of n iid $Ber(p)$ random variables.

Ex2

Suppose $X_1 \sim Bin(n_1, p)$ and $X_2 \sim Bin(n_2, p)$. Let $Y = X_1 + X_2$. Find the distribution of Y.

$$
\begin{aligned}
\psi_Y(t) &= \psi_{X_1}(t)\psi_{X_2}(t) \\
&= (1 - p + pe^t)^{n_1} \cdot (1 - p + pe^t)^{n_2} \\
&= (1 - p + pe^t)^{n_1 + n_2}
\end{aligned}
$$

Note that $(1 - p + pe^t)^{n_1+n_2}$ is the MGF of $Bin(n_1 + n_2, p)$. Thus we know $Y \sim Bin(n_1 + n_2, p)$.

Central Limit Theorem

Motivation Consider the sample mean again and $n = 1000$. Assume $Var(X) = 1$. So far, we have Chebyshev inequality to help us get

$$\Pr(|\bar{X} - \mu| \leq 0.1) \geq 1 - \frac{\sigma^2}{n \cdot (0.01)} = 0.90$$

The lower bound can be crude. For example, if the distribution is $N(\mu, 1)$, then Chebyshev inequality tell us

$$\Pr(|\bar{X} - \mu| \leq 0.1) \geq 1 - \frac{1}{1000 \cdot (0.01)} = 0.90$$

However, we know $\bar{X} \sim N(\mu, \frac{1}{1000})$, hence

$$\sqrt{1000}(\bar{X} - \mu) \sim N(0, 1) \quad \text{and}$$

$$\begin{aligned} & \Pr(|\bar{X} - \mu| \leq 0.1) \\ = \ & \Pr(|\sqrt{1000}(\bar{X} - \mu)| \leq \sqrt{1000} \cdot 0.1) \\ = \ & \Pr(-\sqrt{10} \leq N(0, 1) \leq \sqrt{10}) \\ = \ & \Phi(\sqrt{10}) - \Phi(-\sqrt{10}) = 0.998. \end{aligned}$$

Motivation, contd'

What if the distribution is not normal? For instance, $X_1, \ldots, X_{1000}$ are iid $Ber(p)$, can we calculate the probability? In principle, we can but it is not as easy as the normal case.

What if the distribution is totally unknown? It is not normal, not Bernoulli, not Exponential, not ....? Then we cannot calculate the EXACT value of $\Pr(|\bar{X} - \mu| \leq 0.1)$.

BUT we can calculate a good approximate value of $\Pr(|\bar{X} - \mu| \leq 0.1)$ and it turns out that the approximate value is 0.998.

Central Limit Theorem: Let $X_1, \ldots, X_n$ be iid random variables with common mean μ and common variance σ^2. Then for each real number z,

$$\lim_{n\to\infty} \text{Pr}(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le z) = \text{Pr}(N(0,1) \le z) = \Phi(z)$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Or we say the random variable $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ converges to $N(0,1)$ in distribution. Write

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \to_d N(0,1).$$

or

$$\sqrt{n}(\bar{X}-\mu) \to_d N(0,\sigma^2).$$

Basically, Central Limit Theorem (CLT) says that the distribution of an average will tend to be Normal as the sample size increases, as long as the distribution from which the average is taken is not pathological in the sense that it has a finite mean and finite variance.

CLT will be used to in stat5102 to construct confidence intervals, to calculate p-values in hypothesis testing etc.

Examples of CLT

Ex1. If $X_1, \ldots, X_n$ are iid $Ber(p)$. $\mu = p$ and $\sigma^2 = p(1-p)$.

$$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \to_d N(0,1).$$

Ex2. If $X_1, \ldots, X_n$ are iid $Unif(0, \theta)$. $\mu = \frac{\theta}{2}$, $\sigma^2 = \frac{\theta^2}{12}$

$$\frac{\sqrt{n}(\bar{X} - \frac{\theta}{2})}{\sqrt{\frac{\theta^2}{12}}} \to_d N(0,1).$$

Explain the CLT

Basic idea: We already know that if X and Y have the same MGF, then they have the same distribution. Hence it is not too difficult to imagine that if two random variables have very "similar" MGFs then their distributions are very "similar" too. Hence we show the MGF of $Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ is very close to the MGF of $N(0,1)$.

The MGF of Z_n is

$$\psi_{Z_n}(t) = E[e^{tZ_n}]$$

The MGF of $N(0,1)$ is $e^{\frac{t^2}{2}}$. So we need to argue

$$\psi_{Z_n}(t) \approx e^{\frac{t^2}{2}}.$$

Or

$$\log(\psi_{Z_n}(t)) \approx \frac{t^2}{2}.$$

Explain the CLT, contd'

Note that

$$
\begin{aligned}
tZ_n &= t\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \\
&= \frac{t\sqrt{n}}{\sigma}\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{t\sqrt{n}\mu}{\sigma} \\
&= \sum_{i=1}^{n} \frac{t}{\sqrt{n}\sigma}(X_i-\mu)
\end{aligned}
$$

$$
\begin{aligned}
\psi_{Z_n}(t) = E[e^{tZ_n}] &= E[\prod_{i=1}^{n} e^{\frac{t}{\sqrt{n}\sigma}(X_i-\mu)}] \\
\text{(independence)} &= \prod_{i=1}^{n} E[e^{\frac{t}{\sqrt{n}\sigma}(X_i-\mu)}] \\
\text{(definition)} &= \prod_{i=1}^{n} \psi_{(X_i-\mu)}(\frac{t}{\sqrt{n}\sigma})
\end{aligned}
$$

$\psi_{(X_i-\mu)}$ is the MGF of $(X_i-\mu)$.

Explain the CLT, contd'

$$\log(\psi_{Z_n}(t)) = \sum_{i=1}^{n} \log(\psi_{(X_i-\mu)}(\frac{t}{\sqrt{n}\sigma}))$$
$$= n\log(\psi_{(X_1-\mu)}(\frac{t}{\sqrt{n}\sigma}))$$

We used the assumption that X_is have the same distribution. When n is large $\frac{t}{\sqrt{n}\sigma}$ is close to zero. Use Taylor's expansion of $\psi_{(X_1-\mu)}(\frac{t}{\sqrt{n}\sigma})$ at zero

$$\psi_{(X_1-\mu)}(\frac{t}{\sqrt{n}\sigma}) = \psi_{(X_1-\mu)}(0) + \psi'_{(X_1-\mu)}(0)\frac{t}{\sqrt{n}\sigma}$$
$$+\frac{1}{2}\psi''_{(X_1-\mu)}(0)(\frac{t}{\sqrt{n}\sigma})^2 + R_n$$

R_n is a small reminder term. We know the following properties of MGF

$$\psi_{(X_1-\mu)}(0) = 1 \quad \psi'_{(X_1-\mu)}(0) = E[(X_1-\mu)] = 0$$

$$\psi''_{(X_1-\mu)}(0) = E[(X_1-\mu)^2] = \sigma^2$$

Explain the CLT, contd'

$$\begin{aligned}\psi_{(X_1-\mu)}(\frac{t}{\sqrt{n}\sigma}) &= 1+\frac{1}{2}(\sigma^2)\frac{t^2}{n\sigma^2}+R_n \\ &= 1+\frac{t^2}{2n}+R_n\end{aligned}$$

R_n is a small term compared with $\frac{t^2}{2n}$.

$$\begin{aligned}& n\log(\psi_{(X_1-\mu)}(\frac{t}{\sqrt{n}\sigma})) \\ = \; & n\log(1+\frac{t^2}{2n}+R_n) \\ \approx \; & n(\frac{t^2}{2n}+\text{higher order terms}) \\ \approx \; & \frac{t^2}{2}\end{aligned}$$

Thus we can claim that when $n \to \infty$

$$\psi_{Z_n}(t) \to e^{\frac{t^2}{2}}.$$

for each t.