Week 4 Expected Value: Moment Generating Functions

Instructor: Prof. Phil Everson Presenter: Zhengfei Li (Alex)

I would like to acknowledge that this handout is created with reference to Introduction to Probability by Blitzstein, J.K., and Hwang, J. (2014) and Mathematical Statistics and Data Analysis by J. A. Rice (2007). I would also like to acknowledge the instructions from Prof. Everson.

Preliminaries

Definition (Moments)**.** Let X be an r.v. For any positive integer n, the nth moment of X is $E(X^n)$ if it exists.

Remark: There are other kinds of moments and you may refer to Blitzstein 6.2.

Definition (Moment Generating Function)**.** The moment generating function (MGF) of an r.v. X is $M(t) = E(e^{tX})$, as a function of t, if this is finite on some open interval $(-a, a)$ containing 0. Otherwise we say the MGF of X does not exist.

Remark: By formula of expectation, discrete r.v. X with PMF $P(x)$ would have MGF:

$$M_X(t) = E(e^{tX}) = \sum_{x=-\infty}^{\infty} e^{tx} P(x) \tag{1}$$

And continuous X with PDF $f(x)$ would have MGF:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \tag{2}$$

Remark: Condition for MGF existence is whether the integral is finite on some open interval $(-a, a)$ containing 0. With this condition, we may realize that not all distributions have MGF (e.g. Cauchy distribution has PDF $f(x) = \frac{1}{\pi(1+x^2)}$). In addition, even if MGF exists, it might not be defined for all $t \in \mathbb{R}$ (e.g. for exponential distribution, $M_X(t) = \frac{\lambda}{\lambda - x}$ for $x < \lambda$).

Probably we don't need to fully understand the following? But we will use the following for granted in problem solving

Definition (Characteristic function)**.** The characteristic function of a random variable X is defined to be

$$\phi_X(t) = E(e^{itX}) \tag{3}$$

In continuous case, it is the Fourier Transform of the PDF

$$\phi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f(x) dx \tag{4}$$

Theorem (Inverse theorem). *According to inverse of Fourier Transform,*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{itx}\overline{\phi_X(t)}dx \tag{5}$$

You may learn more from Wikipedia Characteristic function Page

(https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory))

Problem 3

Part a

Please refer to the preliminary section for the definition of MGF and conditions for existence. Below, we are going to assume that we have a continuous random variable X with PDF $f(x) : x \in \mathbb{R}$, unless otherwise specified.

Want to show that $M_X(0) = 1$.

$$\begin{aligned} M_X(0) &= E(e^{0\cdot X}) \\ &= E(1) \\ &= 1 \end{aligned}$$

Want to show that the kth derivative of $M_X(t)$ at $t = 0$ is $E(X^k)$, the kth moment of the distribution (Blitzstein 6.4).

We can start with (2):

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx}f(x)dx \\ M_X'(t) &= \frac{d}{dt}\left[\int_{-\infty}^{\infty} e^{tx}f(x)dx\right] \\ &= \int_{-\infty}^{\infty} \frac{d}{dt}\left[e^{tx}f(x)\right]dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dt}\left[e^{tx}\right]\cdot f(x)dx \\ &= \int_{-\infty}^{\infty} e^{tx}\cdot x\cdot f(x)dx \\ \Rightarrow M_X'(0) &= \int_{-\infty}^{\infty} 1\cdot x\cdot f(x)dx = E(X) \end{aligned}$$

$$
\begin{aligned}
M_X''(t) &= \frac{d}{dt} M_X'(t) \\
&= \frac{d}{dt}\left[\int_{-\infty}^{\infty} e^{tx} \cdot x \cdot f(x)dx\right] \\
&= \int_{-\infty}^{\infty} \frac{d}{dt}\left[e^{tx} \cdot x \cdot f(x)\right] dx \\
&= \int_{-\infty}^{\infty} \frac{d}{dt}\left[e^{tx}\right] \cdot x \cdot f(x)dx \\
&= \int_{-\infty}^{\infty} e^{tx} \cdot x^2 \cdot f(x)dx \\
\Rightarrow M_X''(0) &= \int_{-\infty}^{\infty} 1 \cdot x^2 \cdot f(x)dx = E(X^2)
\end{aligned}
$$

Likewise, we may realize that $\frac{d}{dt}tx = x$ (the chain rule part) is independent of t. Therefore, each time we calculate one higher order derivative of $M_X(t)$, we will multiply the integrand by x. We can then show (by induction) that:

$$
\begin{aligned}
M_X^{(n)}(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot x^n \cdot f(x)dx \\
M_X^{(n)}(0) &= \int_{-\infty}^{\infty} 1 \cdot x^n \cdot f(x)dx = E(X^n)
\end{aligned}
$$

Remark: According to the formula of Taylor expansion:

$$
M_X(t) = \sum_{n=0}^{\infty} M_X^{(n)}(0)\frac{t^n}{n!} = \sum_{n=0}^{\infty} E(X^n)\frac{t^n}{n!}
$$

MGF determines the distribution

We may see the mapping from PDF to MGF as a transformation:

$$
f_X(x) : \mathbb{R} \to \mathbb{R} \xrightarrow{T} M_X(t) : \mathbb{R} \to \mathbb{R}, M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx
$$

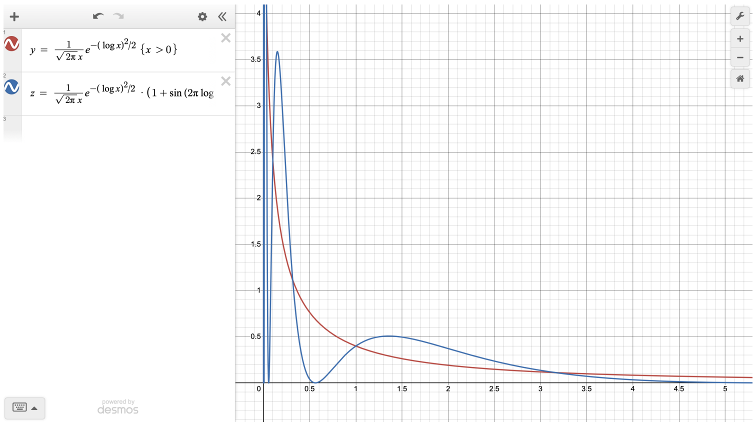
We may compare it with MGF with the characteristic function (4), and we may notice that

$$
M_X(t) = \phi_X(-it)
$$

As we know that Fourier Transform is invertible and we may uniquely map characteristic function to PDF by Inverse theorem (5), we may also uniquely map MGF to PDF. Therefore, we say that MDF determines the distribution.

Remark: "Knowing all the moments of a Random Variable" is a weaker condition than "knowing MGF". This is because the knowledge of MGF also includes information like the existence of MGF. Counterexample: $f_{X_1}(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, x > 0$ and $f_{X_2}(x) = f_{X_1}(x)\left[1 + \sin(2\pi \log x)\right], x > 0$. Please refer to this slide for more details.

https://pages.stat.wisc.edu/ shao/stat609/stat609-05.pdf



Part b

MGF of a linear transformation of a variable.

Recall that normal distribution has PDF:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Let X be an r.v. with MGF $M_X(t)$ and $Y = a + bX$.

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{t(a+bX)}) \\ &= E(e^{at} \cdot e^{tbX}) \\ &= e^{at} E(e^{(bt)X}) \\ &= e^{at} M_X(bt) \end{aligned}$$

Find the MGF for $Z \sim N(0,1)$ and for $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + zt} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2} dz \\ &= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz && \text{integral by recognition} \\ &= e^{\frac{1}{2}t^2} \end{aligned}$$

$$\begin{aligned} M_X(t) &= M_{\mu+z\sigma}(t) \\ &= e^{\mu t} M_Z(\sigma t) \\ &= e^{\mu t} e^{\frac{1}{2}(\sigma t)^2} \end{aligned}$$

Part c

Show that for independent random variables $X_1, ..., X_n$ with MGF's $M_{X_i}(t)$, the MGF for $Y = \sum X_i$ is $M_Y(t) = \prod M_{X_i}(t)$.

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{t(\sum X_i)}) \\ &= E(\prod e^{tX_i}) \\ &= \prod E(e^{tX_i}) && \text{because } X_i\text{'s are independent} \\ &= \prod M_{X_i}(t) \end{aligned}$$

Use this result to show that the sum of independent Normal variables is also Normal

In this question, let's assume that r.v. $X_i \sim N(\mu_i, \sigma_i^2)$ and $Y = \sum X_i$.

$$\begin{aligned} M_Y(t) &= \prod M_{X_i}(t) \\ &= \prod e^{\mu_i t} e^{\frac{1}{2}(\sigma_i t)^2} \\ &= e^{t \sum \mu_i} e^{\frac{t^2}{2} \sum (\sigma_i)^2} \end{aligned}$$

Let's take $\mu_Y = \sum \mu_i$ and $\sigma_Y = \sqrt{\sum (\sigma_i)^2}$

$$\begin{aligned} M_Y(t) &= e^{t \sum \mu_i} e^{\frac{t^2}{2} \sum (\sigma_i)^2} \\ &= e^{t\mu_Y} e^{\frac{1}{2}(\sigma_Y t)^2} \end{aligned}$$

As MGF determines distribution, $Y \sim N(\mu_Y, \sigma_Y^2) = N(\sum \mu_i, \sum (\sigma_i)^2)$

For iid Normal variables, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

As X_i's are iid, $\mu_1 = \mu_2 = \cdots = \mu_n = \mu$ and $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$.

Therefore,

$$\begin{aligned} M_Y(t) &= e^{tn\mu} e^{\frac{t^2}{2} n\sigma^2} \\ M_{\bar{X}}(t) &= M_{\frac{Y}{n}}(t) \\ &= M_Y\left(\frac{y}{n}\right) \\ &= e^{\frac{t}{n} n\mu} e^{\frac{(t/n)^2}{2} n\sigma^2} \\ &= e^{t\mu} e^{\frac{1}{2}((\sigma/\sqrt{n})t)^2} \end{aligned}$$

As MGF determines distribution, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Part d

Theorem (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent random variables having mean 0 and variance σ^2 and the common distribution function F and moment-generating function M defined in a neighborhood of zero. Let*

$$S_n = \sum_{i=1}^{n} X_i$$

Then

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x), x \in \mathbb{R}$$

Sketched proof of Central Limit Theorem

Proof. Let's denote $Z_n = \frac{S_n}{\sigma\sqrt{n}}$, and we notice that Z_n is a linear transformation of S_n. Showing the theorem above is equivalent to showing that $Z_n \dot{\sim} N(0,1)$.

According to our random variable setting, we know the first moment $E(X) = 0$ and second moment $E(X^2) = V(X) + E(X)^2 = \sigma^2$

By the MGF of iid random variables sum:

$$M_{S_n}(t) = [M(t)]^n$$

Using MGF linear transformation

$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

We may use MGF expansion

$$\begin{aligned} M_{Z_n}(t) &= \left[1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^3 \cdot \mathcal{P}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \\ &= \left[1 + \frac{t^2}{2n} + \left(\frac{t}{\sigma\sqrt{n}}\right)^3 \cdot \mathcal{P}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \end{aligned}$$

Using the helpful fact that if $a_n \to a$, then $\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a$, and that $\lim_{n\to\infty}\left(\frac{t}{\sigma\sqrt{n}}\right) = 0$

$$\begin{aligned} \lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \left(\frac{t}{\sigma\sqrt{n}}\right)^3 \cdot \mathcal{P}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \\ &= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n}\right]^n \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Thus, we arrived at the conclusion. □

Remark: With the theorem above, we can then show the more general form of CLT through a linear tranform of S_n that shifts from a zero mean to a non-zero mean.

Remark: Note that CLT is an approximation when we "only" care about sample mean and sample standard deviation for large sample size. For linear combination (e.g.) of i.i.d. normal r.v., it's exactly normal for all n and it's unnecessary to invoke CLT.