MATH 170B LECTURE NOTE 4: TRANSFORMS

HANBAEK LYU

4. Understanding RVs via calculus and analysis

In this section, we will see how we associate a function $M_X(t)$ to each RV X and how we can understand X by looking at $M_X(t)$ instead. The advantage is that now we can use powerful tools from calculus and analysis (e.g., differentiation, integral, power series, Taylor expansion, etc.) to study RVs.

4.1. **Moment generating function.** Let X be a RV. Consider a new RV $g(X) = e^{tX}$, where t is a real-valued parameter we inserted for a reason to be clear soon. A classic point of view of studying X is to look at its *moment generating function* (MGF), which is the expectation $\mathbb{E}[e^{tX}]$ of the RV e^{tX}.

Example 4.1. Let X be a discrete RV with PMF

$$\mathbb{P}(X = x) = \begin{cases} 1/2 & \text{if } x = 2 \\ 1/3 & \text{if } x = 3 \\ 1/6 & \text{if } x = 5. \end{cases} \tag{1}$$

Its MGF is

$$\mathbb{E}[e^{tX}] = \frac{e^{2t}}{2} + \frac{e^{3t}}{3} + \frac{e^{5t}}{6}. \tag{2}$$

Here is a heuristic for why we might be interested in the MGF of X. Recall the Taylor expansion of the exponential function e^s:

$$e^s = 1 + \frac{s}{1!} + \frac{s^2}{2!} + \frac{s^3}{3!} + \cdots. \tag{3}$$

Plug in $s = tX$ and get

$$e^{tX} = 1 + \frac{X}{1!}t + \frac{X^2}{2!}t^2 + \frac{X^3}{3!}t^3 + \cdots. \tag{4}$$

Taking expectation and using its 'linearity', this gives us

$$\mathbb{E}[e^{tX}] = 1 + \frac{\mathbb{E}[X]}{1!}t + \frac{\mathbb{E}[X^2]}{2!}t^2 + \frac{\mathbb{E}[X^3]}{3!}t^3 + \cdots. \tag{5}$$

Notice that the right hand side is a power series in variable t, and the kth moment $\mathbb{E}[X^k]$ of X shows up in the coefficient of the kth order term t^k. In other words, by simply taking the expectation of e^{tX}, we can get all higher moments of X. In this sense, the MGF $\mathbb{E}[e^{tX}]$ generates all moments of X, hence we call its name 'moment generating function'.

As you might have noticed, the equation (5) needs more justification. For example, what if $\mathbb{E}[X^3]$ is infinity? Also, can we really use linearity of expectation for a sum of infinitely many RVs as in the right hand side of (4)? We will get to this theoretical point later, and for now let's get ourselves more familiar to MGF computation.

Example 4.2 (Bernoulli RV)**.** Let $X \sim \text{Bernoulli}(p)$. Then

$$\mathbb{E}[e^{tX}] = e^t p + e^0(1-p) = 1 - p + e^t p. \tag{6}$$

Example 4.3 (Poisson RV)**.** Let $X \sim \text{Poisson}(\lambda)$. Then using the Taylor expansion of the exponential function,

$$\mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{kt}\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^t\lambda)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}. \tag{7}$$

Exercise 4.4 (Geometric RV)**.** Let $X \sim \text{Geom}(p)$. Show that

$$\mathbb{E}[e^{tX}] = \frac{pe^t}{1-(1-p)e^t}. \tag{8}$$

Example 4.5 (Uniform RV)**.** Let $X \sim \text{Uniform}([a,b])$. Then

$$\mathbb{E}[e^{tX}] = \int_a^b e^{tx}\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}. \tag{9}$$

Example 4.6 (Exponential RV)**.** Let $X \sim \text{Exp}(\lambda)$. Then

$$\mathbb{E}[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,dx = \lambda\int_0^\infty e^{(t-\lambda)x}\,dx. \tag{10}$$

Considering two cases when $t<\lambda$ and $t \ge \lambda$, we get

$$\mathbb{E}[e^{tX}] = \begin{cases} \frac{\lambda}{\lambda-t} & \text{if } t<\lambda \\ \infty & \text{if } t\ge\lambda. \end{cases} \tag{11}$$

Example 4.7 (Standard normal RV)**.** Let $X \sim N(0,1)$. Then

$$\mathbb{E}[e^{tX}] = \int_{-\infty}^\infty e^{tx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-x^2/2+tx}\,dx. \tag{12}$$

By completing square, we can write

$$-\frac{x^2}{2}+tx = -\frac{1}{2}(x^2-2tx) = \frac{1}{2}(x-t)^2+\frac{t^2}{2}. \tag{13}$$

So we get

$$\mathbb{E}[e^{tX}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-(x-t)^2/2}e^{t^2/2}\,dx = e^{t^2/2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-(x-t)^2/2}\,dx. \tag{14}$$

Notice that the integrand in the last expression is the PDF of a normal RV with distribution $N(-t,1)$. Hence the last integral equals 1, so we conclude

$$\mathbb{E}[e^{tX}] = e^{t^2/2}. \tag{15}$$

Exercise 4.8 (MGF of linear transform)**.** Let X be a RV and a,b be constants. Let $M_X(t)$ be the MGF of X. Then show that

$$\mathbb{E}[e^{t(aX+b)}] = e^{bt}M_X(at). \tag{16}$$

Exercise 4.9 (Standard normal)**.** Let $X \sim N(\mu,\sigma^2)$ and $Z \sim N(0,1)$. Using the fact that $\mathbb{E}[e^{tZ}] = e^{t^2/2}$ and Exercise 4.9, show that

$$\mathbb{E}[e^{tY}] = e^{\sigma^2t^2/2+t\mu}. \tag{17}$$

4.2. **Two important theorems about MGFs.** The power series expansion (5) of MGF may not be valid in general. The following theorem gives a sufficient condition for which such an expansion is true. We omit its proof in this lecture.

Theorem 4.10. *Let X be a RV. Suppose there exists a constant $h>0$ such that $\mathbb{E}[e^{tX}]<\infty$ for all $x\in(-h,h)$. Then the kth moment $\mathbb{E}[X^k]$ exists for all $k\ge 0$ and there exists a constant $\varepsilon>0$ such that for all $t\in(-\varepsilon,\varepsilon)$,*

$$\mathbb{E}[e^{tX}] = \sum_{k=0}^\infty \frac{\mathbb{E}[X^k]}{k!}t^k. \tag{18}$$

For each RV X, we say its MGF *exists* whenever the hypothesis of the above theorem holds. One of the consequence of the above theorem is that we can access its kth moment by taking kth derivative of its MGF and evaluating at $t=0$.

Exercise 4.11. Suppose the MGF of a RV X exists. Then show that for each integer $k\geq 0$,

$$\frac{d^k}{dt^k}\mathbb{E}[e^{tX}]\Big|_{t=0}=\mathbb{E}[X^k]. \tag{19}$$

Example 4.12 (Poisson RV)**.** Let $X\sim\text{Poisson}(\lambda)$. In Example 4.3, we have computed

$$\mathbb{E}[e^{tX}]=e^{\lambda(e^t-1)}\qquad\forall t\in\mathbb{R}. \tag{20}$$

Differentiating by t and evaluating at $t=0$, we get

$$\mathbb{E}[X]=\frac{d}{dt}e^{\lambda(e^t-1)}\Big|_{t=0}=e^{\lambda(e^t-1)}\lambda e^t\Big|_{t=0}=\lambda. \tag{21}$$

We can also compute its second moment as

$$\mathbb{E}[X^2]=\frac{d^2}{dt^2}e^{\lambda(e^t-1)}\Big|_{t=0}=\frac{d}{dt}\lambda e^{\lambda(e^t-1)+t}\Big|_{t=0}=\lambda e^{\lambda(e^t-1)+t}(\lambda e^t+1)\Big|_{t=0}=\lambda(\lambda+1). \tag{22}$$

This also implies that

$$\text{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\lambda(\lambda+1)-\lambda^2=\lambda. \tag{23}$$

Example 4.13 (Exponential RV)**.** Let $X\sim\text{Exp}(\lambda)$. Our calculation in Example 4.6 implies that

$$\mathbb{E}[e^{tX}]=\frac{\lambda}{\lambda-t}\qquad t\in(-\lambda,\lambda). \tag{24}$$

We can compute the first and second moment of X:

$$\mathbb{E}[X]=\frac{d}{dt}\frac{\lambda}{\lambda-t}\Big|_{t=0}=\frac{\lambda}{(\lambda-t)^2}\Big|_{t=0}=\frac{1}{\lambda} \tag{25}$$

$$\mathbb{E}[X^2]=\frac{d^2}{dt^2}\frac{\lambda}{\lambda-t}\Big|_{t=0}=\frac{d}{dt}\frac{\lambda}{(\lambda-t)^2}\Big|_{t=0}=\frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0}=\frac{2}{\lambda^2}. \tag{26}$$

In fact, by recognizing $\lambda/(\lambda-t)$ as a geometric series,

$$\mathbb{E}[e^{tX}]=\frac{1}{1-t/\lambda}=1+(t/\lambda)+(t/\lambda)^2+(t/\lambda)^3+\cdots \tag{27}$$

$$=1+\frac{1!/\lambda}{1!}t+\frac{2!/\lambda^2}{2!}t^2+\frac{3!/\lambda^3}{3!}t^3+\cdots. \tag{28}$$

Hence by comparing with (18), we conclude that $\mathbb{E}[X^k]=k!/\lambda^k$ for all $k\geq 0$.

The second theorem for MGFs is that they determine the distribution of RVs. This will be critically used later in the proof of the central limit theorem.

Theorem 4.14. *Let X,Y, and X_n for $n\geq 1$ be RVs whose MGFs exist.*

(i) *(Uniqueness) Suppose $\mathbb{E}[e^{tX}]=\mathbb{E}[e^{tY}]$ for all sufficiently small t. Then $\mathbb{P}(X\leq s)=\mathbb{P}(Y\leq s)$ for all $s\in\mathbb{R}$.*

(ii) *(Continuity) Suppose $\lim_{n\to\infty}\mathbb{E}[e^{tX_n}]=\mathbb{E}[e^{tX}]$ for all sufficiently small t and that $\mathbb{E}[e^{tX}]$ is continuous at $t=0$. Then $\mathbb{P}(X_n\leq s)\to\mathbb{P}(X\leq s)$ for all s such that $\mathbb{P}(X\leq x)$ is continuous at $x=s$.*

4.3. **MGF of sum of independent RVs.** One of the nice properties of MGFs is the following factorization for sums of independent RVs.

Proposition 4.15. *Let X,Y be independent RVs. Then*

$$\mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}]. \tag{29}$$

If you believe that the RVs e^{tX} and e^{tY} are independent, then the proof of the above result is one-line:

$$\mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}e^{tY}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}]. \tag{30}$$

In general, it is a special case of the following result.

Proposition 4.16. *Let X,Y be independent RVs. Then for any integrable functions $g_1,g_2:\mathbb{R}\to\mathbb{R}$, we have*

$$\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]. \tag{31}$$

Proof. If X,Y are continuous RVs,

$$\begin{aligned}
\mathbb{E}[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_{X,Y}(x,y)\,dx\,dy && (32)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y)\,dx\,dy && (33)\\
&= \int_{-\infty}^{\infty} g_1(x)f_X(x)\left(\int_{-\infty}^{\infty} g_2(y)f_Y(y)\right)dy && (34)\\
&= \mathbb{E}[g_2(Y)]\int_{-\infty}^{\infty} g_1(x)f_X(x)\,dx && (35)\\
&= \mathbb{E}[g_1(X)]\mathbb{E}[g_2(Y)]. && (36)
\end{aligned}$$

For discrete RVs, use summation and PMF instead of integral and PDF. □

Exercise 4.17 (Binomial RV)**.** Let $X\sim\text{Binomial}(n,p)$. Use the MGF of Bernoulli RV and Proposition 4.15 to show that

$$\mathbb{E}[e^{tX}] = (1-p+e^t p)^n. \tag{37}$$

Example 4.18 (Sum of independent Poisson RVs)**.** Let $X_1\sim\text{Poisson}(\lambda_1)$ and $X_2\sim\text{Poisson}(\lambda_2)$ be independent Poisson RVs. Let $Y = X_1+X_2$. Using Exercise 4.3, we have

$$\mathbb{E}[e^{tY}] = \mathbb{E}[e^{tX_1}]\mathbb{E}[e^{tX_2}] = e^{(\lambda_1+\lambda_2)(e^t-1)}. \tag{38}$$

Notice that the last expression is the MGF of a Poisson RV with rate $\lambda_1+\lambda_2$. By the Uniqueness of MGF (Theorem 4.14 (i)), we conclude that $Y\sim\text{Poisson}(\lambda_1+\lambda_2)$.

Exercise 4.19 (Sum of independent normal RVs)**.** Let $X_1\sim N(\mu_1,\sigma_1^2)$ and $X_2\sim N(\mu_2,\sigma_2^2)$ be independent normal RVs.

(i) Show that $\mathbb{E}[e^{t(X_1+X_2)}] = \exp[(\sigma_1^2+\sigma_2^2)t^2/2 + t(\mu_1+\mu_2)]$.
(ii) Conclude that $X_1+X_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

4.4. **Sum of random number of independent RVs.** Suppose $X_1,X_2,\cdots$ are independent and identically distributed (i.i.d.) RVs and let N be another independent RV taking values in nonnegative integers (e.g., Binomial). For a new RV Y by

$$Y = X_1+X_2+\cdots+X_N. \tag{39}$$

Note that we are summing a random number of X_i's, so there are two sources of randomness that determines Y. As usual, we use conditioning to study such RVs. For instance,

$$\mathbb{E}[Y\,|\,N=n] = \mathbb{E}[X_1+\cdots+X_n] = \mathbb{E}[X_1]+\cdots+\mathbb{E}[X_n] = n\mathbb{E}X_1 \tag{40}$$

$$\mathrm{Var}(Y \mid N = n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\mathrm{Var}(X_1). \tag{41}$$

Hence iterated expectation gives

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|N]] = \mathbb{E}[N\mathbb{E}[X_1]] = \mathbb{E}[X_1]\mathbb{E}[N]. \tag{42}$$

On other other hand, law of total variance gives

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid N)] + \mathrm{Var}(\mathbb{E}[Y \mid N]) \tag{43}$$
$$= \mathbb{E}[N\mathrm{Var}(X_1)] + \mathrm{Var}(N\mathbb{E}[X_1]) \tag{44}$$
$$= \mathrm{Var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2\mathrm{Var}(N). \tag{45}$$

Furthermore, can we also figure out the MGF of Y? After all, MGF is an expectation so we can also get it by iterated expectation. First we compute the conditional version. Denoting $M_X(t) = \mathbb{E}[e^{tX}]$,

$$\mathbb{E}[e^{tY} \mid N = n] = \mathbb{E}[e^{t(X_1+\cdots+X_n)}] = \mathbb{E}[e^{tX_1} \cdots e^{tX_n}] \tag{46}$$
$$= \mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}] = \mathbb{E}[e^{tX_1}]^n = M_{X_1}(t)^n \tag{47}$$
$$= e^{n\log M_{X_1}(t)}. \tag{48}$$

The last line is the trick here. Now the iterated expectation gives

$$\mathbb{E}[e^{tY}] = \mathbb{E}[\mathbb{E}[e^{tY} \mid N]] = \mathbb{E}[e^{(\log M_{X_1}(t))N}]. \tag{49}$$

Note that the last expression is nothing but the MFG of N evaluated at $\log M_{X_1}(t)$ instead of t. Hence

$$\mathbb{E}[e^{tY}] = M_N(\log M_{X_1}(t)). \tag{50}$$

Let us summarize what have obtained so far.

Proposition 4.20. *Let $X_1, X_2, \cdots$ be i.i.d. RVs and let N be another independent RV which takes values from nonnegative integers. Let $Y = \sum_{k=0}^{N} X_k$. Denote the MGF of any RV Z by $M_Z(t)$. Then we have*

$$\mathbb{E}[Y] = \mathbb{E}[N]\mathbb{E}[X_[1]] \tag{51}$$
$$\mathrm{Var}[Y] = \mathrm{Var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2\mathrm{Var}(N) \tag{52}$$
$$M_Y(t) = M_N(\log M_{X_1}(t)). \tag{53}$$

Example 4.21. Let $X_i \sim \mathrm{Exp}(\lambda)$ for $i \geq 0$ and let $N \sim \mathrm{Poisson}(\nu)$. Suppose all RVs are independent. Define $Y = \sum_{k=1}^{N} X_i$. Then

$$\mathbb{E}[Y] = \mathbb{E}[N]\mathbb{E}[X_1] = \lambda/\lambda = 1, \tag{54}$$
$$\mathrm{Var}(Y) = \mathrm{Var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2\mathrm{Var}(N) = \frac{\lambda}{\lambda^2} + \frac{\lambda}{\lambda^2} = \frac{2}{\lambda}. \tag{55}$$

On other hand, recall that $M_{X_1}(t) = \frac{\lambda}{\lambda - t}$ and $M_N(t) = e^{\lambda(e^t-1)}$. Hence

$$\mathbb{E}[e^{tY}] = e^{\lambda(\exp\log\frac{\lambda}{\lambda-t}-1)} = e^{\lambda(\frac{\lambda}{\lambda-t}-1)} = e^{\frac{\lambda t}{\lambda-t}}. \tag{56}$$

So we know everything about Y. Knowing the MGF of Y, we could get all the moments of Y. For instance,

$$\mathbb{E}[Y] = \frac{d}{dt}e^{\frac{\lambda t}{\lambda-t}}\Big|_{t=0} = e^{\frac{\lambda t}{\lambda-t}}\frac{\lambda(\lambda-t)+\lambda t}{(\lambda-t)^2}\Big|_{t=0} = 1. \tag{57}$$

Exercise 4.22. Let $X_1, X_2, \cdots$ be i.i.d. RVs and let N be another independent RV which takes values from nonnegative integers. Let $Y = \sum_{k=0}^{N} X_k$. Denote the MGF of any RV Z by $M_Z(t)$. Using the fact that $M_Y(t) = M_N(\log M_{X_1}(t))$, derive

$$\mathbb{E}[Y] = \mathbb{E}[N]\mathbb{E}[X_1], \tag{58}$$

$$\text{Var}[Y] = \text{Var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2\,\text{Var}(N). \tag{59}$$

Example 4.23. Let $X_i \sim \text{Exp}(\lambda)$ for $i \geq 0$ and let $N \sim \text{Geom}(p)$. Let $Y = \sum_{k=1}^{N} X_k$. Suppose all RVs are independent. Recall that

$$M_{X_1}(t) = \frac{\lambda}{\lambda - t}, \qquad M_N(t) = \frac{pe^t}{1-(1-p)e^t}. \tag{60}$$

Hence

$$M_Y(t) = \frac{p\frac{\lambda}{\lambda-t}}{1-(1-p)\frac{\lambda}{\lambda-t}} = \frac{p\lambda}{(\lambda-t)-\lambda(1-p)} = \frac{p\lambda}{p\lambda - t}. \tag{61}$$

Notice that this is the MGF of an $\text{Exp}(p\lambda)$ variable. Thus by uniqueness, we conclude that $Y \sim \text{Exp}(p\lambda)$. If you remember, sum of k independent $\text{Exp}(\lambda)$ RVs were not an exponential RV (its distribution is $\text{Erlang}(k,\lambda)$. See Exercise 1.19 in Note 1). But as we have seen in this example, if you sum a random number of independent exponentials, they could by exponential again.

Exercise 4.24. Let $X_i \sim \text{Geom}(q)$ for $i \geq 0$ and let $N \sim \text{Geom}(p)$. Suppose all RVs are independent. Let $Y = \sum_{k=0}^{N} X_k$.

(i) Show that the MGF of Y is given by

$$\mathbb{E}[e^{tY}] = \frac{pqe^t}{1-(1-pq)e^t}. \tag{62}$$

(ii) Conclude that $Y \sim \text{Geom}(pq)$.