

## TRANSFORMS AND MOMENT GENERATING FUNCTIONS

Maurice J. Dupré  
Department of Mathematics  
New Orleans, LA 70118  
email: mdupre@tulane.edu  
November 2010

There are several operations that can be done on functions to produce new functions in ways that can aid in various calculations and that are referred to as transforms. The general setup is a transform operator  $\mathcal{T}$  which is really a function whose domain is in fact a set of functions and whose range is also a set of functions. You are used to thinking of a function as something that gives a numerical output from a numerical input. Here, the domain of  $\mathcal{T}$  is actually a set  $\mathcal{D}$  of functions and for  $f \in \mathcal{D}$  the output of  $\mathcal{T}$  is denoted  $\mathcal{T}(f)$  and is in fact a completely new function. A simple example would be differentiation, as the differentiation operator  $D$  when applied to a differentiable function gives a new function which we call the derivative of the function. Likewise, antidifferentiation gives a new function when applied to a function. As another example, we have the Laplace transform,  $\mathcal{L}$ , which is defined by

$$[\mathcal{L}(f)](s) = \int_0^{\infty} e^{-sx} f(x) dx.$$

With a general transform, it is best to use a different symbol for the independent variable of the transformed function than that used for the independent variable of the original function in applications, so as to avoid confusion. The idea of the Laplace transform is that it has useful properties for solving differential equations—it turns them into polynomial equations which can then be solved by algebraic methods. A related transform is the Fourier transform,  $\mathcal{F}$ , which is defined by

$$[\mathcal{F}(f)](t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

More generally, we can take any function  $h$  of two variables, say  $w = h(s, x)$  and for any function  $f$  viewed as a function of  $x$ , we can integrate  $h(s, x)f(x)$  with respect to  $x$  and the result is a function of  $s$  alone. Thus we can define the transform  $\mathcal{H}$  by the rule

$$[\mathcal{H}(f)](s) = \int_{-\infty}^{\infty} h(s, x)f(x) dx.$$

A transform like  $\mathcal{H}$  is called an **Integral Transform** as it primarily works through integration. In this situation, we refer to  $h$  as the **Kernel Function** of the transform. Thus, for the Laplace transform  $\mathcal{L}$  and the Fourier transform  $\mathcal{F}$ , the kernel function is an exponential function. Keep in mind that integration usually makes functions smoother. For instance antidifferentiation certainly increases differentiability.

In general, to make good use of a transform you have to know its properties. A very useful property which the preceding examples are easily seen to have is **Linearity**:

$$\mathcal{T}(a \cdot f \pm b \cdot g) = a \cdot \mathcal{T}(f) \pm b \cdot \mathcal{T}(g), \quad a, b \in \mathbb{R}, \quad f, g \in \mathcal{D}.$$

A transform which has the property of being linear is called a linear transform, and we see that for a linear transform, to find the transform of a function which is itself a sum of functions we just work out the transform of each term and add up the results. Notice that the Fourier and

Laplace transforms are linear as integration is linear and differentiation and antidifferentiation are also linear. As an example of a transform which is not linear, we can define

$$[\mathcal{T}(f)](t) = \int_{-\infty}^{\infty} e^{tf(x)} dx.$$

The transform we want to look at in more detail is one that is useful for probability theory. If the Laplace transform is modified so that the limits of integration go from  $-\infty$  to  $\infty$ , the result is called the bilateral Laplace transform,  $\mathcal{L}_b$ , given by

$$[\mathcal{L}_b(f)](s) = \int_{-\infty}^{\infty} e^{-sx} f(x) dx,$$

and there is clearly no real advantage to the negative sign in the exponential if we are integrating over all of  $\mathbb{R}$ . If we drop this negative sign, then we get the reflected bilateral Laplace transform which we can denote simply by  $\mathcal{M}$ , so

$$[\mathcal{M}(f)](t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

This transform has many of the properties of the Laplace transform and its domain is the set of functions which decrease fast enough at infinity so that the integral is defined. For instance, if  $f(x) = e^{-x^2}$ , then  $\mathcal{M}(f)$  is defined or if there is a finite interval for which  $f$  is zero everywhere outside that interval. In this last case, we say that  $f$  has **compact support**.

We see right away that  $\mathcal{M}$  is linear, but we can also notice that if we differentiate  $\mathcal{M}(f)$  with respect to  $t$ , then we have

$$\left[\frac{d}{dt}\mathcal{M}(f)\right](t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} x e^{tx} f(x) dx = [\mathcal{M}(xf)](t).$$

Notice that this tells us that  $\mathcal{M}$  is turning multiplication by  $x$  into differentiation with respect to  $t$ . On the other hand, if we use integration by parts, we find that

$$\left[\mathcal{M}\left(\frac{d}{dx}f\right)\right](t) = \int_{-\infty}^{\infty} e^{tx} f'(x) dx = e^{tx} f(x) \Big|_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} t e^{tx} f(x) dx,$$

and as we are assuming that  $f$  vanishes faster than  $e^{xt}$  as we go to  $\pm\infty$ , it follows that the first term on the right is zero and the second term is  $t[\mathcal{M}(f)](t)$ , so we have

$$\left[\mathcal{M}\left(\frac{d}{dx}f\right)\right](t) = -t[\mathcal{M}(f)](t).$$

If we had used the bilateral Laplace transform instead of  $\mathcal{M}$ , we would not have had this minus sign, and that is one of the real reasons for the negative sign in exponential of the Laplace transform. In any case, for our transform, we see that multiplication by  $x$  gets turned into differentiation with respect to  $t$  and differentiation with respect to  $x$  gets turned into multiplication by  $-t$ .

A simple example is the indicator of an interval  $[a, b]$  which we denote by  $I_{[a,b]}$ . This function is 1 if  $a \leq x \leq b$  and zero otherwise. Thus,

$$[\mathcal{M}(I_{[a,b]})](t) = \int_{-\infty}^{\infty} e^{tx} I_{[a,b]} dx = \int_a^b e^{tx} dx = \frac{e^{tx}}{t} \Big|_{x=a}^{x=b} = \frac{e^{bt} - e^{at}}{t},$$

so finally,

$$[\mathcal{M}(I_{[a,b]})](t) = \frac{e^{bt} - e^{at}}{t}.$$

This is more useful than it looks at first, because if we need the transform of  $x \cdot I_{[a,b]}$  we just differentiate the previous result with respect to  $t$ .

Thus, using the product rule for differentiation,

$$\begin{aligned} [\mathcal{M}(x \cdot I_{[a,b]})](t) &= \frac{d}{dt}([e^{bt} - e^{at}]t^{-1}) = [be^{bt} - ae^{at}]t^{-1} + [e^{bt} - e^{at}](-t^{-2}) \\ &= \frac{1}{t}[be^{bt} - ae^{at}] - \frac{1}{t^2}[e^{bt} - e^{at}]. \end{aligned}$$

Likewise, if we wanted to find  $\mathcal{M}(x^2 \cdot I_{[a,b]})$  we would just differentiate again. Notice that the graph of  $x \cdot I_{[a,b]}$  is just a straight line segment connecting the point  $(a, a)$  to the point  $(b, b)$ . If we have a straight line connecting  $(a, h_a)$  to  $(b, h_b)$ , then it is the graph of the function  $f$  where

$$f = [h_a + m \cdot (x - a)] \cdot I_{[a,b]},$$

where  $m$  is the slope of the line,

$$m = \frac{h_b - h_a}{b - a}.$$

Therefore to find  $\mathcal{M}(f)$ , we just use linearity and the fact that we already know  $\mathcal{M}(I_{[a,b]})$  and  $\mathcal{M}(x \cdot I_{[a,b]})$ . Specifically, we have

$$[\mathcal{M}(f)](t) = (h_b - m \cdot h_a)[\mathcal{M}(I_{[a,b]})](t) + m[\mathcal{M}(x \cdot I_{[a,b]})](t).$$

Another useful property of  $\mathcal{M}$  is the way it handles shifting along the line. If  $g(x) = f(x - c)$ , we know that the graph of  $g$  is just the graph of  $f$  shifted  $c$  units to the right along the horizontal axis. Then

$$[\mathcal{M}(g)](t) = \int_{-\infty}^{\infty} e^{tx} f(x - c) dx,$$

and if we substitute  $u = x - c$ , we have  $dx = du$  whereas  $x = u + c$ , and we have

$$[\mathcal{M}(g)](t) = \int_{-\infty}^{\infty} e^{t(u+c)} f(u) du = e^{ct} \int_{-\infty}^{\infty} e^{tu} f(u) du = e^{ct} [\mathcal{M}(f)](t),$$

so finally here

$$[\mathcal{M}(f(x - c))](t) = e^{ct} [\mathcal{M}(f)](t).$$

Put another way, we can notice that shifting actually defines a transform. For  $c \in \mathbb{R}$ , we can define the transform  $\mathcal{S}_c$  called the **Shift Operator** by

$$[\mathcal{S}_c(f)](x) = f(x - c).$$

Then our previous result says simply

$$\mathcal{M}(\mathcal{S}_c(f)) = e^{ct} \cdot \mathcal{M}(f),$$

or shifting by  $c$  is turned into multiplication by  $e^c$ . For any function  $g$  we can also define the transform  $\mathcal{K}_g$  called the **Multiplication operator** by

$$\mathcal{K}_g(f) = g \cdot f,$$

so

$$[\mathcal{K}(f)](x) = g(x)f(x).$$

We can then say that

$$\mathcal{M}(\mathcal{S}_c(f)) = \mathcal{K}_{e^{ct}}(\mathcal{M}(f)),$$

or

$$\mathcal{M}\mathcal{S}_c = \mathcal{K}_{e^{ct}}\mathcal{M}.$$

The most important property of the Laplace transform, the Fourier transform and our transform  $\mathcal{M}$  is that if we have two functions  $f$  and  $h$  and if these two functions have the same transform, then they are in fact the same function. Better still, the inverse transform can be found, so there is a transform  $\mathcal{R}$ , which reverses  $\mathcal{M}$ , meaning that if

$$F = \mathcal{M}(f),$$

then

$$\mathcal{R}(F) = f.$$

The reverse transform is a little more complicated—it involves complex numbers and calculus with complex numbers, but it is computable. However, the fact that it exists, is the useful fact, since if we recognize the transform, then we can be sure where it came from.

In probability theory, if  $X$  is an unknown or random variable, then we define its **Moment Generating Function**, denoted  $m_X$  by the formula

$$m_X(t) = E(e^{tX}).$$

If we think of the expectation as an integral, we recognize that it is related to our non-linear transform example. However, the useful fact about the moment generating function is that

$$\frac{d}{dt}m_X(t) = E\left(\frac{d}{dt}e^{tX}\right) = E(X \cdot e^{tX}),$$

and likewise

$$\frac{d^n}{dt^n}m_X(t) = E(X^n \cdot e^{tX}).$$

Evaluating these expressions at  $t = 0$  thus gives all the moments  $E(X^n)$  for all non-negative integer powers  $n$ . In particular,  $m_X(0) = E(X^0) = E(1) = 1$ . It is convenient to denote the  $n^{\text{th}}$  derivative of the function  $f$  by  $f^{(n)}$ , so

$$\frac{d^n}{dt^n}m_X = m_X^{(n)},$$

and

$$E(X^n) = m_X^{(n)}(0), \quad n \geq 0.$$

Thus knowledge of the moment generating function allows us to calculate all the moments of  $X$ . The distribution of  $X$  is for practical purposes determined when its **Cumulative Distribution Function** or **cdf** is known. In general, we denote the cdf for  $X$  by  $F_X$ , so

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R},$$

$$0 \leq F_X(x) \leq 1,$$

$$\lim_{x \rightarrow -\infty} F_X(x) = 0,$$

and

$$\lim_{x \rightarrow \infty} F_X(x) = 1.$$

Notice we then have generally

$$P(a < x \leq b) = F_X(b) - F_X(a), \quad a < b.$$

If  $X$  is a continuous random variable, then its distribution is governed by its **Probability Density Function** or (pdf) which is the derivative of its cumulative distribution function (cdf). We denote its pdf by  $f_X$ , so

$$f_X = \frac{d}{dx}F_X.$$

We then have by **The Fundamental Theorem of Calculus**

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b \frac{d}{dx}F_X(x)dx = \int_a^b f_X(x)dx.$$

This means that in terms of the pdf, probability is area under the graph between the given limits. As a consequence we also have

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f_X(x)dx,$$

for any "sufficiently reasonable" function  $h$ . In particular, we have

$$m_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}f_X(x)dx = [\mathcal{M}(f_X)](t),$$

so

$$m_X = \mathcal{M}(f_X).$$

That is, the moment generating function of  $X$  is the transform of its pdf. Since the transform is reversible, this means that if  $X$  and  $Y$  are continuous random variables and if  $m_X = m_Y$ , then  $f_X = f_Y$  and therefore  $F_X = F_Y$ . If  $X$  is a discrete unknown with only a finite number of possible values, then  $m_X(t)$  is simply a polynomial in  $e^t$  so its distribution is easily seen to be determined by its moment generating function. Thus in fact, for any two unknowns  $X$  and  $Y$ , if  $m_X = m_Y$ , then  $X$  and  $Y$  have the same distribution, that is,  $F_X = F_Y$ .

The most important pdf is the normal density. If  $Z$  is standard normal (and therefore with mean zero and standard deviation one), then its pdf is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}.$$

Then

$$\begin{aligned} m_Z(t) &= [\mathcal{M}(f_Z)](t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz}e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2tz)} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2}t^2} e^{-\frac{1}{2}(z-t)^2} dz = e^{\frac{1}{2}t^2}, \end{aligned}$$

so finally we have the simple result for  $Z$  a standard normal:

$$m_Z(t) = e^{\frac{1}{2}t^2}.$$

Just as it is useful to know the properties of the transform  $\mathcal{M}$  it is also useful to realize how simple modifications to  $X$  cause changes in  $m_X$ . For instance, if  $c \in \mathbb{R}$ , then

$$m_{[X \pm c]}(t) = E(e^{t(X \pm c)}) = E(e^{\pm ct}e^{tX}) = e^{\pm ct}E(e^{tX}) = e^{\pm ct}m_X(t).$$

Thus in general,

$$m_{[X \pm c]}(t) = e^{\pm ct} \cdot m_X(t).$$

On the other hand,

$$m_{cX}(t) = E(e^{tcX}) = E(e^{(ct)X}) = m_X(ct),$$

so we always have

$$m_{cX}(t) = m_X(ct).$$

Lets recall that for any unknown  $X$  it is often useful to standardize in terms of standard deviation units from the mean. That is, we define the new unknown  $Z_X$ , given by

$$Z_X = \frac{X - \mu_X}{\sigma_X},$$

so

$$X = \mu + \sigma \cdot Z_X.$$

We call  $Z_X$  the **Standardization** of  $X$ . This means

$$m_X(t) = e^{\mu t} m_Z(\sigma_X t), \quad Z = Z_X, \quad \mu = \mu_X.$$

Now, if  $X$  is normal with mean  $\mu$  and standard deviation  $\sigma$ , then its standardization,  $Z = Z_X$ , is a standard normal unknown and

$$m_X(t) = e^{\mu t} \cdot m_{\sigma Z}(t) = e^{\mu t} \cdot m_Z(\sigma t) = e^{t\mu} \cdot e^{\frac{1}{2}(\sigma t)^2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

This means that the moment generating function of the normal random variable  $X$  is

$$m_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad \mu = \mu_X, \quad \sigma = \sigma_X.$$

When we get big complicated expressions up in the exponent, it is usually better to adopt the notation  $\exp$  for the exponential function with base  $e$ , so

$$\exp(x) = e^x, \quad x \in \mathbb{R},$$

and for  $X$  any normal random variable,

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

Going the other way, if we know that moment generating function  $m_X$ , for the unknown  $X$ , then as  $Z_X = (X - \mu_X)/\sigma_X$ , we have

$$m_Z(t) = m_{(X-\mu)}\left(\frac{t}{\sigma}\right) = \exp\left(-\frac{\mu t}{\sigma}\right) \cdot m_X\left(\frac{t}{\sigma}\right).$$

Lets see what we can deduce about the distribution of  $X^2$  when we know that for  $X$ . First of all, we know that  $X^2 \geq 0$ , and if  $t \geq 0$ , then

$$F_{X^2}(t) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}),$$

so

$$F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}), \quad x \geq 0$$

and

$$F_{X^2}(x) = 0, \quad x < 0.$$

If  $X$  is continuous, then so is  $X^2$  and the pdf for  $X^2$  can be found by differentiating the cdf using the Chain Rule:

$$f_{X^2}(x) = \frac{d}{dx} F_{X^2}(x) = \frac{d}{dx} [F_X(\sqrt{x}) - F_X(-\sqrt{x})] = f_X(\sqrt{x}) \cdot \frac{1}{2} t^{-1/2} - f_X(-\sqrt{x}) \frac{(-1)}{2} x^{-1/2},$$

so

$$f_{X^2}(x) = \frac{1}{2}(f_X(\sqrt{x}) + f_X(-\sqrt{x})) \cdot \frac{1}{\sqrt{x}}, \quad x \geq 0, \quad f_{X^2}(x) = 0, \quad x < 0.$$

We say that the function  $g$  is symmetric about the origin if  $g(-x) = g(x)$ , always. For instance if  $X$  is normal with mean  $\mu = 0$ , then the pdf for  $X$  is symmetric about the origin. If  $X$  has a pdf symmetric about the origin, then our formula simplifies to

$$f_{X^2}(x) = \frac{f_X(\sqrt{x})}{\sqrt{x}}, \quad x \geq 0, \quad f_{X^2}(x) = 0, \quad x < 0.$$

In particular, if  $Z$  is standard normal, then using the pdf for  $Z$  we find

$$f_{Z^2}(z) = \frac{f_Z(\sqrt{z})}{\sqrt{z}} = \frac{\exp(-z/2)}{\sqrt{(2\pi z)}}, \quad z \geq 0,$$

and of course,

$$f_{Z^2}(z) = 0, \quad z < 0.$$

From the distribution for  $Z^2$  for standard normal  $Z$ , we can now calculate the moment generating function for  $Z^2$ . Thus

$$m_{Z^2}(t) = \int_{-\infty}^{\infty} e^{tz} f_{Z^2}(z) dz = \int_0^{\infty} e^{tz} \frac{\exp(-z/2)}{\sqrt{(2\pi z)}} dz = \frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} e^{tz} \frac{e^{-z/2}}{\sqrt{z}} dz.$$

Here we can substitute  $u = \sqrt{z}$ , so  $z = u^2$  and therefore  $dz = 2udu$ . When we do, we find (don't forget to check the limits of integration-they stay the same here, why?)

$$m_{Z^2}(t) = \frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} \exp(-(1-2t)u^2/2) \frac{1}{u} 2udu = \frac{2}{\sqrt{(2\pi)}} \int_0^{\infty} \exp(-(1-2t)u^2/2) du.$$

If we put

$$\sigma_t = \frac{1}{\sqrt{1-2t}},$$

then

$$2t - 1 = \frac{1}{\sigma_t^2},$$

so our integrand becomes

$$\exp(-(1-2t)u^2/2) = \exp\left(-\frac{1}{2} \cdot \left(\frac{u}{\sigma_t}\right)^2\right),$$

and we recognize this as proportional to the pdf for a normal random variable  $U$  with mean 0 and standard deviation  $\sigma_t$ . Therefore we can see what the integral is easily by looking at it as

$$m_{Z^2}(t) = (2 \cdot \sigma_t) \frac{1}{\sigma_t \sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{1}{2} \cdot \left(\frac{u}{\sigma_t}\right)^2\right) du.$$

But, since the distribution of  $U$  is symmetric about zero,  $P(U \geq 0) = 1/2$ , and therefore

$$\frac{1}{\sigma_t \sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{1}{2} \cdot \left(\frac{u}{\sigma_t}\right)^2\right) du = P(U \geq 0) = \frac{1}{2}.$$

Putting this into the expression for the generating function of  $Z^2$  and cancelling now gives

$$m_{Z^2}(t) = \sigma_t = \frac{1}{\sqrt{1-2t}}.$$

If we want the moments of the unknown  $X$ , it is often laborious to calculate all the derivatives of  $m_X$  and evaluate at zero. In addition, sometimes the expression for the moment generating function has powers of  $t$  in denominators, as for instance in case  $X$  has uniform distribution in the interval  $[a, b]$  since then

$$m_X(t) = \frac{e^{bt} - e^{at}}{(b-a) \cdot t}.$$

Differentiating with the quotient rule would lead to ever more complicated expressions the more times we differentiate. We must keep in mind that as long as the pdf is vanishing sufficiently fast at infinity, which is certainly the case if it has compact support, then  $m_X$  is analytic and can be expressed as the power series in a neighborhood of zero. We then have

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

It is easily seen, if you accept that power series can be differentiated termwise, that for any power series

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n,$$

we have

$$c_n = \frac{f^{(n)}(a)}{n!}, \quad n \geq 0.$$

This means that if we can calculate all the derivatives at  $x = a$ , then we can write down the power series but also it means that if we have the power series we can instantly write down all the derivatives at  $x = a$ , using

$$f^{(n)}(a) = c_n \cdot n!.$$

For instance in case of determining  $m_{Z^2}$ , we can note that as

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad x \in \mathbb{R},$$

it follows that

$$m_{Z^2}(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n \cdot n!},$$

and therefore all odd power moments are zero and for the even powers,

$$E((Z^2)^n) = E(Z^{2n}) = (2n)! \frac{1}{2^n \cdot n!} = \frac{(2n)!}{2^n \cdot n!},$$

thus giving all the moments of  $Z^2$  as well. Therefore the power series expansion of  $m_{Z^2}$  for standard normal  $Z$  is

$$m_{Z^2}(t) = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n \cdot n!} \frac{t^n}{n!} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n \cdot (n!)^2} t^n.$$

The same applies to calculating moment generating functions using  $\mathcal{M}$  because even though we have

$$\mathcal{M}(I_{[a,b]}) = \frac{e^{bt} - e^{at}}{t},$$

this function is analytic and if we write out the power series, the  $t$  cancels out of the denominator, so then when we calculate  $\mathcal{M}(x \cdot I_{[a,b]})$  we do not have to deal with the quotient rule for differentiation, we just differentiate the power series termwise. Thus,



$$\mathcal{M}(I_{[a,b]}) = \frac{1}{t} \left[ \sum_{n=0}^{\infty} \frac{(bt)^n}{n!} - \sum_{n=0}^{\infty} \frac{(at)^n}{n!} \right] = \frac{1}{t} \sum_{n=0}^{\infty} \frac{(b^n - a^n)}{n!} t^n = \sum_{n=1}^{\infty} \frac{(b^n - a^n)}{n!} t^{n-1}.$$

Thus we have

$$\mathcal{M}(I_{[a,b]}) = (b - a) + \frac{b^2 - a^2}{2} t + \frac{b^3 - a^3}{3 \cdot 2} t^2 + \sum_{n=4}^{\infty} \frac{(b^n - a^n)}{n!} t^{n-1}.$$

To calculate  $\mathcal{M}(x \cdot I_{[a,b]})$  we just differentiate termwise, so

$$\mathcal{M}(x \cdot I_{[a,b]}) = \frac{d}{dt} \sum_{n=1}^{\infty} \frac{(b^n - a^n)}{n!} t^{n-1} = \sum_{n=2}^{\infty} \frac{(b^n - a^n)}{n!} (n-1) t^{n-2},$$

or

$$\mathcal{M}(x \cdot I_{[a,b]}) = \sum_{n=2}^{\infty} \frac{(b^n - a^n)}{n!} (n-1) t^{n-2} = \frac{b^2 - a^2}{2} + \frac{b^3 - a^3}{3} t + \sum_{n=4}^{\infty} \frac{(b^n - a^n)}{n!} (n-1) t^{n-2},$$

and using this together with the power series for the indicator itself we can calculate the moment generating function for any unknown whose pdf is a straight line segment supported on  $[a, b]$ .

We can also break up the transform process over disjoint intervals. Thus, if  $f$  and  $g$  are functions with domain  $[a, b]$  and  $[c, d]$  respectively where  $b \leq c$ , then

$$h = f \cdot I_{[a,b]} + g \cdot I_{[c,d]},$$

is the function which agrees with  $f$  on  $[a, b]$  and agrees with  $g$  on  $[c, d]$ , and using linearity of  $\mathcal{M}$  we have

$$\mathcal{M}(h) = \mathcal{M}(f \cdot I_{[a,b]}) + \mathcal{M}(g \cdot I_{[c,d]}).$$

In particular, we can easily find the transform of any function whose graph consists of straight line segments such as the graph of a frequency polygon or a histogram.

One of the most useful facts about the moment generating function is that if  $X$  and  $Y$  are independent, then  $m_{(X+Y)} = m_X \cdot m_Y$ . This is because if  $X$  and  $Y$  are independent, then  $f(X)$  and  $g(Y)$  are uncorrelated for any functions  $f$  and  $g$ , so

$$E(f(X) \cdot g(Y)) = [E(f(X))] \cdot [E(g(Y))],$$

and in particular then

$$E(e^{tX} \cdot e^{tY}) = [E(e^{tX})] \cdot [E(e^{tY})]$$

so

$$m_{(X+Y)}(t) = E(e^{t(X+Y)}) = E(e^{tX} \cdot e^{tY}) = [E(e^{tX})] \cdot [E(e^{tY})] = m_X(t) \cdot m_Y(t).$$

In particular, for any independent random sample  $X_1, X_2, X_3, \dots, X_n$  of the random variable  $X$ , we have

$$F_{X_k} = F_X, \quad k = 1, 2, 3, \dots, n,$$

that is to say, all have the same distribution as  $X$ , and therefore

$$m_{X_k} = m_X, \quad k = 1, 2, 3, \dots, n.$$

But this means that for the sample total  $T_n$  given by

$$T_n = \sum_{k=1}^n X_k = X_1 + X_2 + X_3 + \dots + X_n,$$

we must have

$$m_{T_n} = m_{X_1} \cdot m_{X_2} \cdot m_{X_3} \cdots m_{X_n} = [m_X]^n,$$

so we have the simple result

$$m_{T_n} = m_X^n,$$

that is the moment generating function for  $T_n$  is simply the result of taking the  $n^{\text{th}}$  power of  $m_X$ . For the sample mean

$$\bar{X}_n = \frac{1}{n} T_n,$$

we then have its moment generating function

$$m_{\bar{X}_n}(t) = m_{T_n}(t/n) = [m_X(t/n)]^n.$$

Finally, we should observe that if  $X$  and  $Y$  are unknowns and if there is  $\delta > 0$  with

$$m_X(t) = m_Y(t), \quad |t| < \delta,$$

then

$$m_X^{(n)}(0) = m_Y^{(n)}(0), \quad n = 0, 1, 2, 3, \dots$$

and therefore

$$E(X^n) = E(Y^n), \quad n = 0, 1, 2, 3, \dots$$

which in turn implies that

$$E(f(X)) = E(f(Y)), \quad f \in \mathcal{P},$$

where  $\mathcal{P}$  denotes the set of all polynomial functions. But in fact, on any bounded closed interval, any continuous bounded function can be uniformly approximated arbitrarily closely by polynomial functions,

and this means that

$$E(f(X)) = E(f(Y)), \quad f \in \mathcal{B},$$

where  $\mathcal{B}$  denotes the set of all bounded continuous functions which have compact support, and that in turn implies that  $F_X = F_Y$ . Thus, if  $m_X$  and  $m_Y$  agree on an open neighborhood of zero then  $X$  and  $Y$  must have the same distribution. In fact, this is most easily seen for the case where  $X$  and  $Y$  are continuous by using an inversion transform for  $\mathcal{M}$  which we can denote by  $\mathcal{R}$ . Thus for continuous  $X$

$$\mathcal{R}(\mathcal{M}(f_X)) = f_X,$$

but

$$m_X = \mathcal{M}(f_X),$$

therefore

$$f_X = \mathcal{R}(\mathcal{M}(f_X)) = \mathcal{R}(m_X).$$

Thus, if  $X$  and  $Y$  are continuous unknowns and if  $m_X = m_Y$ , then

$$f_X = \mathcal{R}(m_X) = \mathcal{R}(m_Y) = f_Y,$$

and therefore  $F_X = F_Y$ , that is, both  $X$  and  $Y$  have the same distribution. In the general case, let  $V(X) \subset \mathbb{R}$  denote the set of possible values of  $X$ . If  $X$  is a simple unknown, then  $V(X)$  is finite and each  $v \in V(X)$  has positive probability and those probabilities add up to one. Then

$$m_X(t) = \sum_{v \in V(X)} [P(X = v)] \cdot e^{tv},$$

the sum being finite as there are only a finite number of these values with positive probability, and therefore we can define

$$p_X(w) = m_X(\ln w) = \sum_{v \in V(X)} [P(X = v)] \cdot w^v,$$

from which we see that  $p_X$  completely encodes the distribution of  $X$  and it is called the characteristic polynomial of  $X$ . Since equality of two polynomials in a neighborhood of one (as  $\log(1) = 0$ ,) implies their equality (their difference would be a polynomial with infinitely many roots and therefore zero), this means that for discrete  $X$  and  $Y$  both simple, if  $m_X = m_Y$ , then  $p_X = p_Y$  and  $P(X = v) = P(Y = v)$ , for every  $v \in \mathbb{R}$ , which is to say they have the same distribution. If  $X$  is a general discrete unknown, then there is a sequence of numbers  $v_1, v_2, v_3, \dots$  with

$$V(X) = \{v_1, v_2, v_3, \dots\},$$

and

$$m_X(t) = \sum_{v \in V(X)} [P(X = v)] \cdot e^{tv} = \sum_{n=1}^{\infty} [P(X = v_n)] \cdot e^{tv_n}.$$

It is reasonable that the infinite sums here are approximated by finite partial sums for which the simple unknown case applies so that for any discrete unknown  $m_X$  determines its distribution. One way to look at this case is to think of  $v \in V(X)$  as a variable but the coefficients (the probabilities) in the summation are held constant, so if we differentiate  $m_X$  with respect to a particular  $v$  in the sum treating all the other values as constants, then we have

$$\frac{\partial}{\partial v} m_X = P(X = v) \cdot t e^{tv},$$

so

$$\frac{\partial^2}{\partial v^2} m_X = P(X = v) \cdot e^{tv} (1 + t^2),$$

and putting  $t = 0$  in this last expression gives the value  $P(X = v)$  as the second derivative with respect to  $v$  when  $t = 0$ , that is

$$P(X = v) = \left( \frac{\partial^2}{\partial v^2} m_X \right)_{t=0}.$$

This means that  $m_X$  determines the probabilities of all possible values of  $X$  which then gives the distribution of  $X$  and the cumulative distribution function  $F_X$  for  $X$ . To be more precise here, if we have

$$f(t) = \sum_{n=0}^{\infty} p_n \cdot e^{v_n t},$$

then for a given  $k$  we can view  $f$  as also depending on a variable  $w$  giving a function of two variables  $f_k(t, w)$  by setting

$$f_k(t, w) = p_k \cdot e^{wt} + \sum_{n \neq k} p_n \cdot e^{v_n t},$$

and then

$$\frac{\partial^2}{\partial w^2} f_k(t, v_k) = p_k.$$

Now admittedly, this is not determining the coefficients from  $f$  as a function of  $t$  alone, but it does seem to make plausible that  $f$  determines all the coefficients  $p_n$ ,  $n = 1, 2, 3, \dots$

DEPARTMENT OF MATHEMATICS, TULANE UNIVERSITY, NEW ORLEANS, LA 70118  
*E-mail address:* `mdupre@tulane.edu`