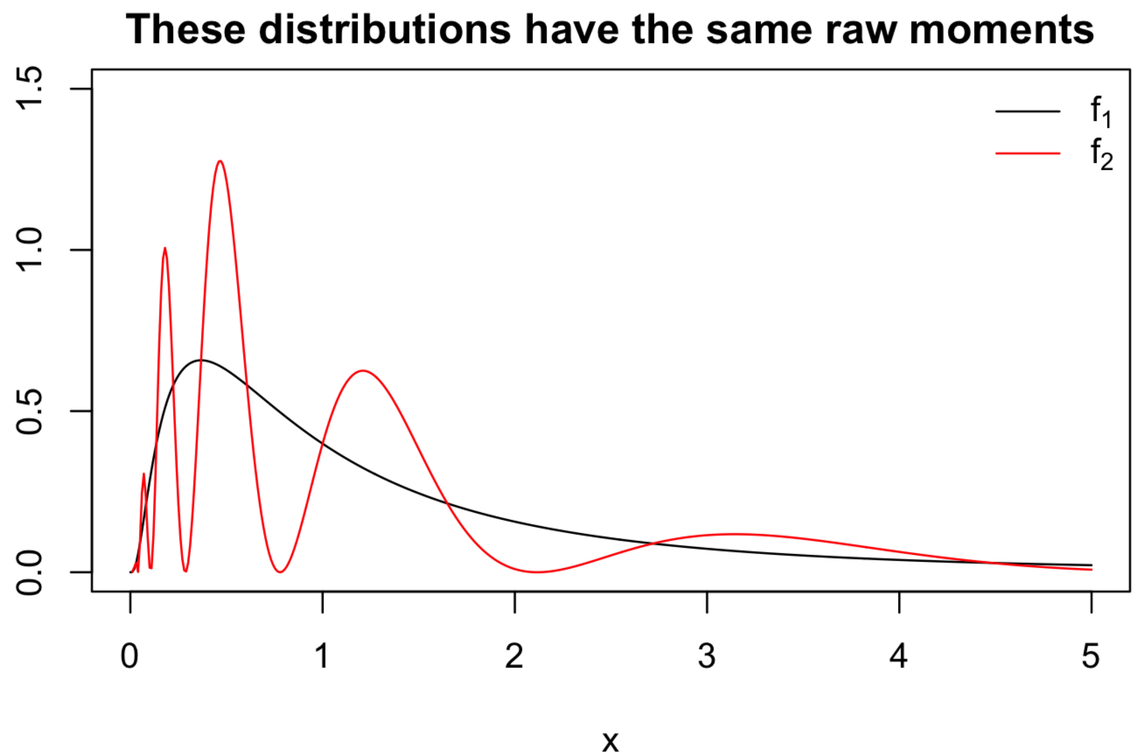

Figure 1.31

1.8 The moment-generating function

We have referred many times to the fact that objects like the cdf, the pmf, or the pdf fully and uniquely characterize the entire distribution of a random variable. If you have any one of these objects, you *have* the distribution, and you know everything there is to know about it. Moments are an example of an object that does *not* fully and uniquely characterize the distribution of a random variable. Two random variables can have the exact same moments, and yet be wildly different.

Example 1.22. Consider $X \sim \mathrm{N}(2,\ 2)$ and $Y \sim \mathrm{Poisson}(2)$. So these distributions are very different (discrete versus continuous, different ranges entirely, etc), and yet $\mathbb{E}[X] = E(Y) = 2$ and $\mathrm{var}(X) = \mathrm{var}(Y) = 2$.

That example shows that the first two moments are not enough to uniquely characterize the distribution, but it turns out that even if *all* of the infinitely-many moments $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\mathbb{E}[X^3]$, ... are the same, the distributions could still be very different:

Example 1.23. Consider two absolutely continuous random variables X_1 and X_2 with pdfs

$$
\begin{aligned}
f_1(x) &= \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{1}{2}(\ln x)^2\right), && x \geq 0 \\
f_2(x) &= f_1(x)[1 + \sin(2\pi \ln x)], && x \geq 0.
\end{aligned}
$$

Figure 1.31 displays these densities, and clearly they are quite different. And yet, you can show that $\mathbb{E}[X_1^n] = \mathbb{E}[X_2^n]$ for all $n \in \mathbb{N}$. We start by computing $\mathbb{E}[X_2^n]$ via LOTUS:

$$
\begin{aligned}
\mathbb{E}[X_2^n] &= \int_0^\infty x^n f_2(x)\,\mathrm{d}x \\
&= \int_0^\infty x^n f_1(x)[1 + \sin(2\pi \ln x)]\,\mathrm{d}x \\
&= \int_0^\infty [x^n f_1(x) + x^n f_1(x)\sin(2\pi \ln x)]\,\mathrm{d}x \\
&= \int_0^\infty x^n f_1(x)\,\mathrm{d}x + \int_0^\infty x^n f_1(x)\sin(2\pi \ln x)]\,\mathrm{d}x \\
&= \mathbb{E}[X_1^n] + \int_0^\infty x^n f_1(x)\sin(2\pi \ln x)]\,\mathrm{d}x \\
&= \mathbb{E}[X_1^n] + \int_0^\infty \frac{x^n}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\ln x)^2\right)\sin(2\pi \ln x)]\,\mathrm{d}x
\end{aligned}
$$

So it will suffice to show that the second term is zero. We apply the change of variables

$$
\begin{aligned}
u &= \ln x - n \\
x &= e^{u+n} \\
\mathrm{d}u &= \frac{1}{x}\,\mathrm{d}x.
\end{aligned}
$$

So

$$
\begin{aligned}
\mathbb{E}[X_2^n] &= \mathbb{E}[X_1^n] + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\left(e^{u+n}\right)^n e^{-(u+n)^2/2}\sin(2\pi(u+n))\,\mathrm{d}u \\
&= \mathbb{E}[X_1^n] + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{nu+n^2} e^{-(u+n)^2/2}\sin(2\pi u + 2\pi n)\,\mathrm{d}u \\
&= \mathbb{E}[X_1^n] + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{nu+n^2} e^{-u^2/2-nu-n^2/2}\sin(2\pi u)\,\mathrm{d}u \\
&= \mathbb{E}[X_1^n] + \int_{-\infty}^\infty \underbrace{\frac{1}{\sqrt{2\pi}} e^{-(u^2-n^2)/2}\sin(2\pi u)}_{\text{odd function, integrates to zero}}\,\mathrm{d}u \\
&= \mathbb{E}[X_1^n].
\end{aligned}
$$

Alright alright. So moments, even infinitely many of them, are not enough to uniquely characterize a distribution. But there exists an object intimately related to the moments that does uniquely characterize its distribution:

Definition 1.24. Let X be any random variable, discrete or continuous. The **moment-generating**

function (**mgf**) of X is

$$M_X(t) = \mathbb{E}[e^{tX}]. \tag{1.30}$$

Remark 1.14. If we plug in $t = 0$, we get $M_X(0) = \mathbb{E}[e^0] = 1$. But for other values $t \neq 0$, the mgf may not be finite. As we will see next, the mgf is a powerful object when it is finite for all t in at least a small interval around zero.

Theorem 1.14. Let X and Y be two random variables such that $M_X(t) < \infty$ and $M_Y(t) < \infty$ for all $-b < t < b$ for some $b > 0$, and furthermore, $M_X(t) = M_Y(t)$ for all $-b < t < b$. Then X and Y have the exact same distribution.

We will not prove this result, but it means that the mgf, like the cdf or pmf or pdf, is one of those objects that fully and uniquely characterizes the entire distribution of a random variable. If you know the mgf, you know everything.

That's all well and good, but what's the point of this new object? Here's the point:

Theorem 1.15. Let X be a random variables such that $M_X(t) < \infty$ for all $-b < t < b$ for some $b > 0$. Then the nth derivative of the mgf evaluated at zero is equal to the nth raw moment:

$$M_X^{(n)}(0) = \mathbb{E}[X^n] \quad \forall n \in \mathbb{N}. \tag{1.31}$$

Proof sketch. We can rewrite M_X with a Taylor series expansion about zero (ie a Maclaurin series):

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

We can also rewrite e^{tX} using the Taylor series expansion of the exponential function

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

While it is not always the case that the expected value of an infinite sum is the infinite sum of the expected values, it is true in this case (for reasons I will not go into), and so

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

By definition, we know that $M_X(t) = \mathbb{E}[e^{tX}]$, and so it must be the case that

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

The only way these can be equal if is each term is equal, and so $M_X^{(n)}(0) = \mathbb{E}[X^n]$. $\square$

Remark 1.15. Theorem 1.15 is killer. It allows us to compute moments by taking a derivative instead of simplifying an integral or an infinite series. As a general matter, differentiation is much more straightforward than integration, so this will be very useful. Now that we have this tool, we can use it to do some moment calculations that we've been putting off.

Remark 1.16. The moment-generating function is intimately related to something called the *Laplace transform*, which you may encounter in a course on ordinary differential equations.

Example 1.24. Let $X \sim \text{Poisson}(\lambda)$, and fix any $t \in \mathbb{R}$. Then

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] \\
&= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=0}^{\infty} \left(e^t\right)^k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda e^t - \lambda}.
\end{aligned}$$

Note that we did not have to place any restrictions on t to complete the computation, so the mgf is defined for all t and we have

$$M_X(t) = \exp\left(\lambda e^t - \lambda\right), \quad t \in \mathbb{R}. \tag{1.32}$$

We can use this to compute the mean and variance of the Poisson:

$$\begin{aligned}
M_X'(t) &= \lambda e^t \exp(\lambda e^t - \lambda) \\
M_X''(t) &= (\lambda e^t)^2 \exp(\lambda e^t - \lambda) + \lambda e^t \exp(\lambda e^t - \lambda)
\end{aligned}$$

$$\begin{aligned}
M_X'(0) &= \mathbb{E}[X] = \lambda \\
M_X''(0) &= E(X^2) = \lambda^2 + \lambda
\end{aligned}$$

So $\mathbb{E}[X] = \lambda$ and $\text{var}(X) = E(X^2) - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. Once you have the mgf, getting these moment calculations done with derivatives is way easier than proceeding from LOTUS and grinding through an infinite series or integral.

Example 1.25. Let $X \sim \text{Exp}(\lambda)$, so its density is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, and $f(x) = 0$ otherwise. Fix any $t \in \mathbb{R}$. Then

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} \mathrm{d}x = \int_0^{\infty} \lambda e^{(t-\lambda)x} \mathrm{d}x.$$

If $t - \lambda \geq 0$, this integral is infinite. If $t - \lambda < 0$, the integral evaluates to $\lambda/(\lambda - t)$. So

$$M_X(t) = \begin{cases} \frac{\lambda}{\lambda - t}, & t < \lambda, \\ \infty, & t \geq \lambda. \end{cases}$$

We can now compute the moments of X by differentiating $M_X(t)$. But for this example, there is an alternative, even easier route to computing the moments. Observe that for $t < \lambda$, we can divide the numerator and denominator by λ to get $M_X(t) = \lambda/(\lambda - t) = 1/(1 - t/\lambda)$. We can now use the geometric series: for any $r \in (-1, 1)$, we have $1/(1 - r) = 1 + r + r^2 + r^3 + \dots$. So as long as we plug in $t \in (-\lambda, \lambda)$, we get $t/\lambda = r \in (-1, 1)$, and therefore

$$M_X(t) = \frac{1}{1 - t/\lambda} = 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \frac{t^3}{\lambda^3} + \dots. \tag{1.33}$$

But recall from the proof sketch for Theorem 1.15, we also have, by definition, the following Taylor series expansion:

$$M_X(t) = M_X(0) + M_X'(0)t + \frac{M_X''(0)}{2!}t^2 + \frac{M_X'''(0)}{3!}t^3 + \dots. \tag{1.34}$$

Matching coefficients in front of t, t^2, t^3, and so on, we have

$$\frac{M_X^{(n)}(0)}{n!} = \frac{1}{\lambda^n}.$$

So

$$\mathbb{E}[X^n] = M_X^{(n)}(0) = \frac{n!}{\lambda^n}.$$

Example 1.26. Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Computing the mgf will be an exercise in completing the square, so let's review that first.

Suppose we have e.g. $3x^2 - 5x$. We want to write it in the form $\square(x - \square)^2 + \square$, where we fill in the $\square$'s with some numbers. This is how we can do that:

1. Factor out the number in front of x^2:
   $3x^2 - 5x = 3\left(x^2 - \frac{5}{3}x\right)$.

2. Inside the parentheses, make a 2 appear in front of the linear-in-x term by multiplying and dividing it by 2:
   $x^2 - \frac{5}{3}x = x^2 - 2 \cdot \frac{5}{6}x$.

3. Recall that $(x - a)^2 = x^2 - 2ax + a^2$. In our expression $x^2 - 2 \cdot \frac{5}{6}x$, we have both the x^2 and the $-2ax$, with $a = 5/6$. So let's add and subtract a^2:

$$x^2 - 2 \cdot \frac{5}{6}x = x^2 - 2 \cdot \frac{5}{6}x + \left(\frac{5}{6}\right)^2 - \left(\frac{5}{6}\right)^2 = \left(x - \frac{5}{6}\right)^2 - \left(\frac{5}{6}\right)^2.$$

4. Put everything together:

$$\begin{aligned} 3x^2 - 5x &= 3\left(x^2 - \frac{5}{3}x\right) = 3\left(x^2 - 2 \cdot \frac{5}{6}x\right) \\ &= 3\left(\left(x - \frac{5}{6}\right)^2 - \left(\frac{5}{6}\right)^2\right) \\ &= 3\left(x - \frac{5}{6}\right)^2 - \frac{25}{12}. \end{aligned}$$

Let's put completing the square into action to compute the normal mgf. Fix any $t \in \mathbb{R}$. We have

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] && \text{definition} \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x && \text{LOTUS} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x && \text{combine base-}e\text{ terms.} \end{aligned} \tag{1.35}$$

Next, we follow the completing-the-square procedure for the quadratic $tx - (x-\mu)^2/(2\sigma^2)$ in the exponent:

1. Factor out the number in front of x^2:
   $tx - \frac{1}{2\sigma^2}(x-\mu)^2 = -\frac{1}{2\sigma^2}(-2\sigma^2 tx + (x-\mu)^2)$, then simplify what's inside the parentheses:

$$\begin{aligned} -2\sigma^2 tx + (x-\mu)^2 &= -2\sigma^2 tx + x^2 - 2\mu x + \mu^2 \\ &= x^2 - 2(\mu + \sigma^2 t)x + \mu^2. \end{aligned}$$

2. Recall that $(x-a)^2 = x^2 - 2ax + a^2$. In our expression $x^2 - 2(\mu+\sigma^2 t)x + \mu^2$, we have both the x^2 and the $-2ax$, with $a = \mu + \sigma^2 t$. So let's add and subtract a^2:

$$\begin{aligned} x^2 - 2(\mu+\sigma^2 t)x + \mu^2 &= x^2 - 2(\mu+\sigma^2 t)x + (\mu+\sigma^2 t)^2 - (\mu+\sigma^2 t)^2 + \mu^2 \\ &= \left(x - \mu - \sigma^2 t\right)^2 - (\mu+\sigma^2 t)^2 + \mu^2 \\ &= \left(x - \mu - \sigma^2 t\right)^2 - 2\mu\sigma^2 t - \sigma^4 t^2. \end{aligned}$$

3. Put everything together:

$$\begin{aligned} tx - \frac{1}{2\sigma^2}(x-\mu)^2 &= -\frac{1}{2\sigma^2}\left(\left(x-\mu-\sigma^2 t\right)^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right) \\ &= -\frac{1}{2\sigma^2}\left(x-\mu-\sigma^2 t\right)^2 + \mu t + \frac{1}{2}\sigma^2 t^2. \end{aligned}$$

Plugging this last expression back into the exponent in (1.35), we get

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{1}{2}\sigma^2 t^2}\,\mathrm{d}x \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}}\,\mathrm{d}x}_{\text{the integral of a Gaussian pdf}} \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

In the last line, we observed that the integral is just the integral of a Gaussian pdf with mean $\mu + \sigma^2 t$ and variance σ^2, and all pdfs integrate to 1. Note two things:

- We did not have to place any restrictions on t to complete the computation, so the mgf is defined for all t;
- This is a classic example of "massage and squint." We just pushed and pulled the integrand and massaged it with constants until it turned into something familiar. Then we applied an identity.

So the mgf of $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \quad t \in \mathbb{R}. \tag{1.36}$$

We can use this to compute the mean and variance of the normal:

$$\begin{aligned} M_X'(t) &= (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \\ M_X''(t) &= (\mu + \sigma^2 t)^2 \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) + \sigma^2 \exp\left(\frac{1}{2}t^2\right) \end{aligned}$$

$$\begin{aligned} M_X'(0) &= \mathbb{E}[X] = \mu \\ M_X''(0) &= \mathbb{E}[X^2] = \mu^2 + \sigma^2. \end{aligned}$$

So $\mathbb{E}[X] = \mu$ and $\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$. Not exactly earth-shattering, but note that it's the first time we've actually seen a calculation of these facts.

Example 1.27 (Using the mgf to determine the distribution of $aX + b$)**.** We first observe a cool fact: if you know the mgf of X, then you can derive the mgf of $aX + b$, without ever having to go through pdfs. Indeed, we compute:

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

Using this, let's prove the fact from (1.23), which is that a shifted and scaled Gaussian is still Gaussian. To do this, let's derive the mgf of $aX + b$ for $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$M_{aX+b}(t) = e^{bt}M_X(at) = e^{bt+\mu(at)+\frac{1}{2}\sigma^2(at)^2} = e^{(b+a\mu)t+\frac{1}{2}\sigma^2a^2t^2}.$$

Here, we plugged in (1.36) to get the second inequality. But the final expression on the righthand side is just the mgf for $Y = \mathcal{N}(a\mu + b, \sigma^2 a^2)$! Since $M_{aX+b} = M_Y$, we know $aX + b$ and Y must have the same distribution, by Theorem 1.14. Therefore $aX + b$ is Gaussian.

Theorem 1.16. If the moment generating function of a random variable X exists, then all of its moments are finite: $E[|X|^n] < \infty$ for all $n \in \mathbb{N}$.

However, the converse is not true. A random variable could have all finite moments, and yet the mgf does not exist. So all moments finite is a *necessary condition* for the existence of the moment generating function, but not a sufficient one.

Summary of basic facts about moments and mgfs

- mgf exists **does imply** $E(X^n)$ finite for all $n \in \mathbb{N}$;
- If mgf exists, then $M_X^{(n)}(0) = E(X^n)$ for all $n \in \mathbb{N}$;
- $E(X^n)$ finite for all $n \in \mathbb{N}$ **does not imply** mgf exists;
- $M_X = M_Y$ **does imply** $F_X = F_Y$;
- $E(X^n) = E(Y^n)$ for all $n \in \mathbb{N}$ **does not imply** $F_X = F_Y$.