4 Moment generating functions

Moment generating functions (mgf) are a very powerful computational tool. They make certain computations much shorter. However, they are only a computational tool. The mgf has no intrinsic meaning.

4.1 Definition and moments

Definition 1. *Let X be a random variable. Its moment generating function is*

$$M_X(t) = \mathbf{E}[e^{tX}]$$

At this point in the course we have only considered discrete RV's. We have not yet defined continuous RV's or their expectation, but when we do the definition of the mgf for a continuous RV will be exactly the same.

Example: Let X be geometric with parameter p. Find its mgf.
Recall that $f_X(k) = p(1-p)^{k-1}$. Then

$$M(t) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = pe^t \sum_{k=1}^{\infty} e^{t(k-1)}(1-p)^{k-1} = pe^t \frac{1}{1-e^t(1-p)}$$

Note that the geometric series that we just summed only converges if $e^t(1-p) < 1$. So the mgf is not defined for all t.

What is the point? Our first application is show that you can get the moments of X from its mgf (hence the name).

Proposition 1. *Let X be a RV with mgf $M_X(t)$. Then*

$$\mathbf{E}[X^n] = M_X^{(n)}(0)$$

where $M_X^{(n)}(t)$ is the nth derivative of $M_X(t)$.

Proof.

$$\frac{d^n}{dt^n}\mathbf{E}[e^{tX}] = \frac{d^n}{dt^n}\sum_k e^{tk} f_X(k) = \sum_k k^n e^{tk} f_X(k)$$

At $t = 0$ this becomes

$$\sum_k k^n f_X(k) = \mathbf{E}[X^n]$$

□

There was a cheat in the proof. We interchanged derivatives and an infinite sum. You can't always do this and to justify doing it in the above computation we need some assumptions on $f_X(k)$. We will not worry about this issue.

Example: Let X be binomial RV with n trials and probability p of success. The mgf is

$$\begin{aligned}
\mathbf{E}[e^{tX}] &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} = [pe^t + (1-p)]^n
\end{aligned}$$

Now we use it to compute the first two moments.

$$\begin{aligned}
M'(t) &= n[pe^t + (1-p)]^{n-1} pe^t, \\
M''(t) &= n(n-1)[pe^t + (1-p)]^{n-2} p^2 e^{2t} + n[pe^t + (1-p)]^{n-1} pe^t
\end{aligned}$$

Setting $t = 0$ we have

$$\mathbf{E}[X] = M'(0) = np, \mathbf{E}[X^2] = M''(0) = n(n-1)p^2 + np$$

So the variance is

$$var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$$

4.2 Sums of independent random variables

Suppose X and Y are independent random variables, and we define a new random variable by $Z = X + Y$. Then the pmf of Z is given by

$$f_Z(z) = \sum_{x,y:x+y=z} f_X(x) f_Y(y)$$

The sum is over all points (x, y) subject to the constraint that they lie on the line $x + y = z$. This is equivalent to summing over all x and setting $y = z - x$. Or we can sum over all y and set $x = z - y$. So

$$f_Z(z) = \sum_x f_X(x) f_Y(z-x), \quad f_Z(z) = \sum_y f_X(z-y) f_Y(y)$$

Note that this formula look like a discrete convolution. One can use this formula to compute the pmf of a sum of independent RV's. But computing the mgf is much easier.

Proposition 2. *Let X and Y be independent random variables. Let $Z = X + Y$. Then the mgf of Z is given by*

$$M_Z(t) = M_X(t) M_Y(t)$$

If $X_1, X_2, \cdots, X_n$ are independent and identically distributed, then

$$M_{X_1+X_2+\cdots+X_n}(t) = [M(t)]^n$$

where $M(t) = M_{X_j}(t)$ is the common mgf of the X_j's.

Proof.

$$\mathbf{E}[e^{tZ}] = \mathbf{E}[e^{t(X+Y)}] = \mathbf{E}[e^{tX}\, e^{tY}] = \mathbf{E}[e^{tX}]\, \mathbf{E}[e^{tY}] = M_X(t) M_Y(t)$$

The proof for n RV's is the same. □

Computing the mgf does not give you the pmf of Z. But if you get a mgf that is already in your catalog, then it effectively does. We will illustrate this idea in some examples.

Example: We use the proposition to give a much shorter computation of the mgf of the binomial. If X is binomial with n trials and probability p of success, then we can write it as a sum of the outcome of each trial:

$$X = \sum_{j=1}^{n} X_j$$

where X_j is 1 if the jth trial is a success and 0 if it is a failure. The X_j are independent and identically distributed. So the mgf of X is that of X_j raised to the n.

$$M_{X_j}(t) = \mathbf{E}[e^{tX_j}] = pe^t + 1 - p$$

So

$$M_X(t) = \left[pe^t + 1 - p\right]^n$$

which is of course the same result we obtained before.

Example: Now suppose X and Y are independent, both are binomial with the same probability of success, p. X has n trials and Y has m trials. We argued before that $Z = X + Y$ should be binomial with $n + m$ trials. Now we can see this from the mgf. The mgf of Z is

$$M_Z(t) = M_X(t)M_Y(t) = \left[pe^t + 1 - p\right]^n \left[pe^t + 1 - p\right]^m = \left[pe^t + 1 - p\right]^{n+m}$$

which is indeed the mgf of a binomial with $n + m$ trials.

Example: Look at the negative binomial distribution. It has two parameters p and n and the pmf is

$$f_X(k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k \geq n$$

So

$$\begin{aligned} M_X(t) &= \sum_{k=n}^{\infty} e^{tk} \binom{k-1}{n-1} p^n (1-p)^{k-n} \\ &= \sum_{k=n}^{\infty} e^{tk} \frac{(k-1)!}{(n-1)!(k-n)!} p^n (1-p)^{k-n} \end{aligned}$$

Let $j = k - n$ in the sum to get

$$\begin{aligned} & \sum_{j=0}^{\infty} e^{t(n+j)} \frac{(n+j-1)!}{(n-1)!j!} p^n (1-p)^j \\ = & \frac{e^{tn}p^n}{(n-1)!} \sum_{j=0}^{\infty} \frac{(n+j-1)!}{j!} e^{tj} (1-p)^j \\ = & \frac{e^{tn}p^n}{(n-1)!} \sum_{j=0}^{\infty} \frac{d^{n-1}}{dx^{n-1}} x^{n+j-1} |_{x=e^t(1-p)} \\ = & \frac{e^{tn}p^n}{(n-1)!} \frac{d^{n-1}}{dx^{n-1}} \sum_{j=0}^{\infty} x^{n+j-1} |_{x=e^t(1-p)} \end{aligned}$$

The natural thing to do next is factor out an x^{n-1} from the series to turn it into a geometric series. We do something different that will save some computation later. Note that the $n-1$th derivative will kill any term x^k with $k < n-1$. So we can replace

$$\sum_{j=0}^{\infty} x^{n+j-1} \quad by \quad \sum_{j=0}^{\infty} x^j$$

in the above. So we have

$$\begin{aligned}
\frac{e^{tn}p^n}{(n-1)!}\frac{d^{n-1}}{dx^{n-1}}\sum_{j=0}^{\infty} x^j|_{x=e^t(1-p)} &= \frac{e^{tn}p^n}{(n-1)!}\frac{d^{n-1}}{dx^{n-1}}\frac{1}{1-x}|_{x=e^t(1-p)} \\
&= \frac{e^{tn}p^n}{(n-1)!}(n-1)!\frac{1}{1-x}|_{x=e^t(1-p)} \\
&= \left[\frac{e^t p}{1-e^t(1-p)}\right]^n
\end{aligned}$$

This is of the form something to the n. The something is just the mgf of the geometric distribution with parameter p. So the sum of n independent geometric random variables with the same p gives the negative binomial with parameters p and n.

4.3 Other generating functions

The book uses the "probability generating function" for random variables taking values in $0, 1, 2, \cdots$ (or a subset thereof). It is defined by

$$G_X(s) = \sum_{k=0}^{\infty} f_X(k) s^k$$

Note that this is just $\mathbf{E}[s^X]$, and this is our mgf $\mathbf{E}[e^{tX}]$ with $t = \ln(s)$. Anything you can do with the probability generating function you can do with the mgf, and we will not use the probability generating function.

The mgf need not be defined for all t. We saw an example of this with the geometric distribution where it was defined only if $e^t(1-p) < 1$, i.e, $t < -\ln(1-p)$. In fact, it need not be defined for any t other than 0. As an example of this consider the RV X that takes on all integer values and $P(X = k) = c(1+k^2)^{-1}$. The constant c is given by

$$\frac{1}{c} = \sum_{k=-\infty}^{\infty} \frac{1}{1+k^2}$$

We leave it to the reader to show that

$$\sum_{k=-\infty}^{\infty} e^{tk} \frac{1}{1+k^2} = \infty$$

for all nonzero t.

Another moment generating function that is used is $\mathbf{E}[e^{itX}]$. A probabilist calls this the *charateristic function* of X. An analyst might call it the fourier transform of the distribution of X. It has the advantage that for real t it is always defined.

End of September 28 lecture
