Moment Generating Functions

May 1, 2020

Recall that if X is a random variable then the moment generating function (MGF) of X is

$$m_X(t) = \mathbb{E}[e^{Xt}]$$

which you compute by multiplying e^{xt} by the pdf/pmf and integrating/summing.

For nice random variables knowing the pdf is equivalent to knowing the MGF, but unfortunately it is not so easy to go from a MGF back to the pdf, except in special cases where we know the answer. In this class we'll exploit the special cases we know to compute the distribution of the sum of independent random variables.

At the end of the last section, we gave a formula for the pdf of $X+Y$, when X and Y are independent random variables. It is a nice formula, but working out the bounds of integration ends up usually being a bit tricky. We will see in a moment that if we know the MGF of X and Y then computing the MGF of $X+Y$ is very easy. If we're lucky enough to recognize the MGF of $X+Y$ as the MGF of a pdf that we know, then we can completely avoid computing the integral.

Here's the main theorem of this section:

THEOREM 0.1. *Let X and Y be independent random variables with MGFs $m_X(t)$ and $m_Y(t)$ then the MGF of $X+Y$ is*

$$m_{X+Y}(t) = m_X(t) m_Y(t)$$

Proof. We start with the definition of the MGF

$$m_{X+Y}(t) = \mathbb{E}[e^{(X+Y)t}] = \mathbb{E}[e^{Xt}e^{Yt}]$$

then we distribute the t and use that the exponential of a sum is the product of the exponentials. Then we use that because X and Y are independent

$$\mathbb{E}[e^{Xt}e^{Yt}] = \mathbb{E}[e^{Xt}]\mathbb{E}[e^{Yt}].$$

This is basically what it means to be independent (for any functions g and h, $\mathbb{E}[g(X)h(y)] = \mathbb{E}[g(X)]\mathbb{E}[h(y)]$). But the last equality is just the product of the MGFs of X and Y. Which is what we want to show. □

Now let's see how we can apply this theore.

EXAMPLE 0.2. *Let X be a normal random variably with mean μ_X and variance σ_X^2 and Y be a normal random variably with mean μ_Y and variance σ_Y^2. Furthermore, let X and Y be independent. Use the moment generating functions to compute the pdf of $X+Y$.*

Solution: We have seen the MGF of X is $m_X(t)=e^{\frac{t^2\sigma_X^2}{2}+\mu_X t}=exp\left(\frac{t^2\sigma_X^2}{2}+\mu_X t\right)$ and $m_Y(t) = exp\left(\frac{t^2\sigma_Y^2}{2}+\mu_Y t\right)$.

The above formula tells us the MGF of $X+Y$ is:

$$m_{X+Y}(t)=exp\left(\frac{t^2\sigma_X^2}{2}+\mu_X t\right)exp\left(\frac{t^2\sigma_Y^2}{2}+\mu_Y t\right)$$

which we simplify to:

$$m_{X+Y}(t)=exp\left(\frac{t^2(\sigma_X^2+\sigma_Y^2)}{2}+(\mu_X+\mu_Y)t\right)$$

mostly just using that the product of exp is the exp of the sums.

Then we recognize that this is the MGF of a normal random variable with mean $\mu_X+\mu_Y$ and variance $\sigma_X^2+\sigma_Y^2$. So this means the pdf of $X+Y$ is

$$f_{X+Y}(t)=\frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}}exp\left(-\frac{(t-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right)$$

Note that using the last section we could instead have computed

$$f_{X+Y}(t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(\sigma_X^2)}}exp\left(-\frac{(x-\mu_X)^2}{2(\sigma_X^2)}\right)\frac{1}{\sqrt{2\pi(\sigma_X^2)}}exp\left(-\frac{(t-x-\mu_Y)^2}{2(\sigma_Y^2)}\right)dx$$

which looks like a lot more work to simplify.

In the above example we started with r.v.'s that we already knew the MGF of and then we combined them we recognized a MGF of a rv that we knew. In general, we can't expect this to happen, but it's worth trying in the cases we see here.

Let's look at another example where we can use the MGF to compute the distribution of a sum.

EXAMPLE 0.3. *Let X be a gamma random variable with parameters α_X and β and let Y be a gamma random variable with parameters α_Y and β. Assume X and Y are independent.*

(a) Compute the MGF X and Y.

(b) Compute the MGF $X+Y$.

(c) What is the distribution of $X+Y$?

Note in this example the gamma random variables have the same β parameter but their α parameters can be different. The sum of gamma random variables with different β values doesn't have a nice description, and cannot be easily identified by its MGF.

Solution:

(a) We'll just compute the MGF of a gamma random variable. Then substituting α_X or α_Y answers the question.

The pdf of a gamma random variable is:

$$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta}$$

for $x>0$ and 0 otherwise so its MGF is

$$m(t)=\int_0^{\infty}e^{tx}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta}\ dx$$

We combine the exponential terms to get:

$$m(t)=\int_0^{\infty}e^{-(1/\beta-t)x}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\ dx$$

and then do the u-sub $u=(1/\beta-t)x$ to get

$$m(t)=(1/\beta-t)^{-\alpha}\int_0^{\infty}e^{u}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}u^{\alpha-1}\ du$$

so then combining $(t-1/\beta)^{-\alpha}$ and β^{α} gives

$$m(t)=(1-\beta t)^{-\alpha}$$

where we have used the definition of the $\Gamma(\alpha)$ to notice the remaining terms were just $\Gamma(\alpha)/\Gamma(\alpha)=1$

(b) From our main theorem we have

$$m_{X+Y}(t)=m_X(t)m_Y(t)=(1-\beta t)^{-\alpha_X}(1-\beta t)^{-\alpha_Y}=(1-\beta t)^{-\alpha_X+\alpha_Y}$$

(c) Matching our solution in part (b) to the the computation in part (a), we see that the distribution of $X+Y$ is a gamma random variable with parameters $\alpha_X+\alpha_Y$ and β.

If you think back to the gamma r.v.'s and remember if $\alpha=1$, then this is an exponential random variable. The exponential random variable is used to model the waiting time for the first arrival in a random process. Then the time from the first arrival to the second arrival is also modeled by an independent exponential random. From the result we just computed, this show that the total waiting time until the 2nd arrival is a gamma random variable with $\alpha=2$. Repeating this idea inductively shows that the total waiting time for the k^{th} arrival is a gamma random variable with $\alpha=k$.

We have already seen that representing a binomial random variable as a sum of n independent Bernoulli random variables makes the computation of it's expectation and variance much easier. Using the generalization of the main theorem to the sums of n independent random variable also gives us an easy way to compute the MGF of Binomial r.v.

EXAMPLE 0.4. *Compute the MGF of Y, a Binomial(n,p) random variable.*

Solution:

Let $X_1,X_2,...,X_n$ be independent Bernoulli random variables with probability of success p. Then $\sum_{i=1}^{n} X_i$ has the same distribution of Y.

The MGF of X_1 is

$$m_{X_1}(t) = \mathbb{E}[e^{tX_1}] = e^{t*0}(1-p) + e^{t*1}(p) = (1-p) + e^t p$$

Then the MGF of Y is

$$m_Y(t) = m_{X_1}(t) m_{X_2}(t) ... m_{X_n}(t) = \Big((1-p) + e^t p\Big)^n$$

where we have used that the all have the same MGF so multiplying them together is the same as raising to the n^{th} power.

One more computation I'll leave for you to think about is the sum of independent Possion r.v. When this r.v. was introduced, we used it to model the number of successes in a given time interval. We then claimed that if the length of the time interval was doubled then the number of successes would still be a Poisson r.v. with twice the mean. What this really meant was the number of successes in the first half of the time interval is a Poisson r.v. and the number of successes in the second half of the time interval is another independent Poisson r.v. and the sum of these two Poisson r.v.s gives a new Poisson r.v. whose mean is the sum of the first two Poisson r.v.s. So you can check that in general the sum two independent Poisson random variables with mean λ_1 and λ_2 is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Finally, a very important distribution is in statistics is the χ^2 random variable. The book does a nice example showing the the χ^2 rv's can be related to $Z_1^2 + Z_2^2 + ... + Z_n^2$, where Z_i's are independent standard normal random variables.