Statistics 3657 : Moment Generating Functions

A useful tool for studying sums of independent random variables is generating functions. In this course we consider moment generating functions.

Definition 1 (Moment Generating Function) *Consider a distribution (with X a r.v. with this distribution). We say this distribution (or X) has moment generating function (mgf) given by*

$$M(t) = E(e^{tX})$$

if there exists some $\delta > 0$ such that $M(t) < \infty$ for $t \in (-\delta, \delta)$.

The set or the domain of M is important. We will need to deal with

$$D = \{t : M(t) < \infty\} \ .$$

In general this will depend on the particular distribution, and in the case of distributions with parameters it will depend on the value of the parameter. The interval $(-\delta, \delta) \subseteq R$.

This definition requires $\delta > 0$ because we will require that M has derivatives at 0. Formally, by differentiating under the expectation sign we get

$$\begin{aligned} M'(t) &= \frac{dM(t)}{dt} \\ &= \frac{d\mathrm{E}(e^{tX})}{dt} \\ &= \mathrm{E}\left(\frac{de^{tX}}{dt}\right) \\ &= \mathrm{E}(Xe^{tX}) \end{aligned}$$

Thus we have $M'(0) = \mathrm{E}(X)$. If we differentiate k times we formally obtain

$$\begin{aligned} M^{(k)}(t) &= \frac{d^k M(t)}{dt^k} \\ &= \frac{d^k \mathrm{E}(e^{tX})}{dt^k} \\ &= \mathrm{E}\left(\frac{d^k e^{tX}}{dt^k}\right) \\ &= \mathrm{E}(X^k e^{tX}) \end{aligned}$$

Upon setting $t = 0$ we obtain $M^{(k)}(0) = \mathrm{E}(X^k)$.

Another consequence of $\delta > 0$ is that the function M has a power series about $t = 0$. Formally this is

$$\begin{aligned} M(t) &= \mathrm{E}\left(e^{tX}\right) \\ &= \mathrm{E}\left(\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right) \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{E}\left(X^k\right) \end{aligned}$$

This mean that the coefficients of this power series about 0 (or the McLaurin series) will yield k-th moments about 0.

The student should review from their introductory calculus course text and notes the notion of radius of convergence. It was also reviewed earlier in our course. A power series can be differentiated term by term for arguments inside the disk of radius equal to the radius of convergence. It is for this reason the moment generating function also has a power series representation and so can be differentiated term by term. A distribution without an infinite number of moments cannot have a moment generating function; for example the student's t distribution with any degrees of freedom.

The function M then generates, via the derivatives or the coefficients of the power series, the moments for the corresponding distribution. Hence it is called the moment generating function.

Remark 1 *If the moment generating function for a distribution exits, then the MGF uniquely determines the distribution.*

In the text Remark 1 is called Property A (p155 Edition 3).

This property says, in other words, that there is a 1 to 1 correspondence between the set of distributions with MGFs and the set of MGFs. For example if we have an MGF of the form for a Gamma distribution, then the distribution of a r.v. with this MGF must be a Gamma distribution. This property is really only useful if one has a table of distributions and their MGFs, otherwise it is not easy to know the distribution if one has just the MGF. When this property works, in particular for sums of independent r.v.s (see below) it is very useful to simplify some calculations. This methodology is useful only in some special particular cases, one of which is Markov chains.

There is an inversion formula that allows one to calculate the corresponding pdf or pmf from an MGF. However this is not so important to us in this course. The inversion formula is actually a special case of the Fourier inversion formula, a calculation that involves complex integration, that is line integrals with complex number arguments.

Aside: The 1 to 1 property discussed above deserve another comment. Let

$$\mathcal{A} = \text{the set of distributions with MGFs}$$

$$\mathcal{B} = \text{set of MGFs}$$

This statement is there is a 1 to 1 mapping between the sets $\mathcal{A}$ and $\mathcal{B}$. In this case the domain and range are sets of functions. This complicates things in terms of writing out the mapping and the inverse mapping.

Except for this complication of a mapping between sets of functions, the notion is the same as we have used in this course many times.

The 1 to 1 mapping property is useful because sometimes it is easier to work with one of the objects (eg pdf) instead of MGF and sometimes it is easier to work with the other (MGF instead of pdf). Of course the notion of *to work with* depends on what calculation we are trying to make.

End of Aside

Remark 2 *Not all distributions have moment generating functions. This is because if a distribution has an MGF then it must have finite moments of all orders.*

For example the Cauchy distribution does not have an MGF because the Cauchy distribution does not have moments.

A t distribution does not have a moment generating function. A $t_{(n)}$ distribution, that is a student's t with n degrees of freedom, has density

$$f(x) = \frac{\Gamma\left((n+1)/2\right)}{\sqrt{n\pi}\Gamma\left(n/2\right)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

Thus by a ratio test (see a first year calculus text)

$$\int_R x^k f(x)dx = const \int_{-\infty}^{\infty} x^k \left(1+\frac{x^2}{n}\right)^{-(n+1)/2} dx$$

exists and is finite if $k-(n+1)<-1$ (that is $k<n$) but not for $k \geq n$.

The student should refer to the text for the multivariate moment generating function. There are also other generating functions, including the probability generating function, the Fourier transform (or characteristic generating function) and the Laplace transform. All are used in a similar general spirit, but with some important differences. They are sometimes used, in connection with an inversion formula, to obtain numerical approximations for certain types of distributions in areas of ruin theory and mathematical finance.

We now consider some examples and other properties.

Example [Binomial]

Suppose that $X \sim Bin(n, p)$. Then

$$\begin{aligned} M(t) &= \mathrm{E}(e^{tX})) \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \left(1 - p + pe^t\right)^n \end{aligned}$$

This is defined for all $t \in R$.

Suppose that $Y_i \sim$ Bernoulli(p). Then Y_i has mgf $M(t) = (1 - p + pe^t)$. Suppose that $Y_1, \ldots, X_n$ are iid Bernoulli(p) r.v.'s. Let $X = Y_1 + \ldots + Y_n$. Then X has mgf

$$M_X(t) = \prod_{i=1}^{n} \left(1 - p + pe^t\right) = \left(1 - p + pe^t\right)^n$$

which is the mgf of a Binomial distribution. Thus $X \sim$ Binom(n, p).

Question for the student : What happens if X_i are independent Bernoulli(p_i) with different p_i for each i. We still obtain the product mgf above, but is it the mgf of a Binomial distribution?

End of Example

Example [Gamma] Suppose that $X \sim$ Gamma(α, λ). Then

$$\begin{aligned} M(t) &= \mathrm{E}(e^{tX})) \\ &= \int_0^\infty e^{tx} \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dx \frac{1}{(\lambda-t)^\alpha} \\ & \quad \text{change variables } y = (\lambda - t)x \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} \\ &= \left(\frac{\lambda}{\lambda - t}\right)^\alpha \end{aligned}$$

Notice this integral is finite only if $\lambda - t > 0$, thus $M(t)$ has domain $t < \lambda$, or equivalently $t \in (-\infty, \lambda)$.

Suppose that X and Y are independent Gamma random variables, with parameters (α_1, λ) and (α_2, λ). Note that the parameter λ is the same for both. Consider the r.v. $T = X + Y$. What is the distribution of T?

This can be obtained by a general transformation method, or by the convolution formula. It can also be obtained by the method of mgf's. The mgf of T is

$$M_T(t) = \mathrm{E}\left(e^{tT}\right)$$

$$
\begin{aligned}
&= \mathrm{E}\left(e^{t(X+Y)}\right) \\
&= \mathrm{E}\left(e^{tX}\right)\mathrm{E}\left(e^{tY}\right) \text{ by independence} \\
&= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1}\left(\frac{\lambda}{\lambda-t}\right)^{\alpha_2} \\
&= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1+\alpha_2} .
\end{aligned}
$$

This is the mgf a Gamma($\alpha_1+\alpha_2, \lambda$) distribution. Since each mgf corresponds to exactly one distribution, therefore $T \sim$ Gamma($\alpha_1+\alpha_2, \lambda$).

End of Example

Let us consider a variation on the Gamma example above. Suppose we had X and Y independent r.v.s with Gamma(α_1, λ_1) and Gamma(α_2, λ_2) distributions respectively. Does the sum have an MGF?

Formally this sum $X+Y$ has $\mathrm{E}\left(e^{t(X+Y)}\right)$, which we call M say,

$$
\begin{aligned}
M(t) &= \mathrm{E}\left(e^{t(X+Y)}\right) \\
&= \mathrm{E}\left(e^{tX}\right)\mathrm{E}\left(e^{tY}\right) \text{ by independence} \\
&= \left(\frac{\lambda_1}{\lambda_1-t}\right)^{\alpha_1}\left(\frac{\lambda_2}{\lambda_2-t}\right)^{\alpha_2}
\end{aligned}
$$

Is M and MGF? We have $M(t)$ is finite provided that $t < \lambda_1$ and $t < \lambda_2$. Thus if we take $\delta = \min(\lambda_1, \lambda_2)$ we see that $\delta > 0$ and $M_{X+Y}(t)$ is finite for all $|t| < \delta$. Thus M is an MGF.

This MGF M does determine the distribution of $X+Y$, according to Property A. Except in the special case $\lambda_1 = \lambda_2$ there is no nice simple parametric distribution to which this corresponds.

Exponential Distribution

Recall that the exponential, λ distribution is also a Gamma($1, \lambda$) distribution. Thus we obtain by applying the above mgf type of calculation that if X_i, $i = 1, \ldots, n$ are iid exponential λ r.v.'s, then

$$
X_1 + \ldots + X_n \sim \text{Gamma}(n, \lambda) .
$$

χ^2 Distribution

What is the pdf of a $\chi^2_{(d)}$ distribution? From the material studied about functions of a single r.v. we found the pdf of a $\chi^2_{(1)}$ is

$$
f_1(x) = \left(\frac{1}{2}\right)^{1/2} x^{-1/2} \frac{1}{\sqrt{\pi}} e^{-x/2} \mathrm{I}_{(0,\infty)}(x)
$$

Upon rewriting this we find

$$
f_1(x) = \left(\frac{1}{2}\right)^{1/2} x^{1/2-1} \frac{1}{\Gamma(1/2)} e^{-x/2} \mathrm{I}_{(0,\infty)}(x)
$$

which is the Gamma(1/2,1/2) pdf.

What is the pdf of the sum of d iid $\chi^2_{(1)}$ r.v.s? This distribution is knows as the $\chi^2_{(d)}$ distribution. Let $X_i, i = 1, \ldots, d$ be iid $\chi^2_{(1)}$ r.v.s and $Y = X_1 + \ldots + X_d$. Then the mgf of Y is

$$M_Y(t) = \left(\left(\frac{1/2}{1/2 - t}\right)^{1/2}\right)^d = \left(\frac{1/2}{1/2 - t}\right)^{d/2}$$

Thus Y has a Gamma(d/2 , 1/2) distribution and hence we have the formula for this density.

Recall that a $\chi^2_{(d)}$ distribution is also the Gamma($\frac{d}{2}, \frac{1}{2}$) distribution. Thus the sum of independent χ^2 random variables also has a χ^2 distribution. Specifically if X_i are independent, and if $X_i \sim \chi^2_{(d_i)}$, and $d = d_1 + d_2 + \ldots + d_n$, then

$$X_1 + \ldots + X_n \sim \text{Gamma}\left(\frac{d}{2}, \frac{1}{2}\right)$$

Thus we find the MGF of the sum of independent χ^2 random variables is of a Gamma form with parameters $\frac{d}{2}, \frac{1}{2}$. Recalling again that the Gamma($\frac{d}{2}, \frac{1}{2}$) distribution is also the $\chi^2_{(d)}$ distribution we see that sum random variable has a $\chi^2_{(d)}$ distribution. Notice we just get a χ^2 distribution with degrees of freedom given the sum of the original degrees of freedom.

Example [Normal]

Suppose that $Z \sim N(0, 1)$. Then Z has mgf

$$\begin{aligned} M(t) &= \mathrm{E}\left(e^{tZ}\right) \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2tz + t^2)} dz e^{\frac{t^2}{2}} \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Notice this is finite for all t.

Suppose that $X \sim N(\mu, \sigma^2)$. We can then calculate the MGF by computing the corresponding integral to calculate $\mathrm{E}(e^{tX})$.

On the other hand we can take advantage of our calculations and some use some properties to our advantage, thereby saving some work in our calculations. From our study of 1 dimensional transformations we have found that X can be represented as $X = \sigma Z + \mu$, where $Z \sim N(0, 1)$. Thus X has mgf

$$\begin{aligned} M_X(t) &= \mathrm{E}\left(e^{t(\sigma Z + \mu)}\right) \\ &= e^{t\mu} \mathrm{E}\left(e^{t\sigma Z}\right) \\ &= e^{t\mu + \frac{\sigma^2 t^2}{2}} \end{aligned}$$

since the expectation in the second last line is the mgf of Z evaluated at the argument $t\sigma$. Again notice that this formula is finite for all real values t, so this is indeed an MGF.

At home show that if X and Y are independent normal random variables, then $X + Y$ has a normal distribution. Contrast this with the calculation using the convolution formula, or by a method of completing the transformation and then calculating the marginal.

End of Example

Example [Random Sums of Random Variables or Compound Distributions]

Another important application of generating functions is to random sums of random variables. Here we will also need to make use of conditional expectation, as we did earlier in our study of conditional expectation and random sums.

In that setting we considered N as a Z_0^+ (non negative integer valued random variable), and X_i, $i \geq 1$ as an iid sequence of random variables independent of N. We now suppose that N and X (a generic X_i) have mgf's, say M_N and M_X.

Consider

$$S = \begin{cases} 0 & \text{if} \quad N = 0 \\ \sum_{i=1}^{n} X_i & \text{if} \quad N = n \geq 1 \end{cases}$$

We next find the mgf of S. For $n \geq 1$

$$\begin{aligned} \mathrm{E}\left(e^{tS} \mid N = n\right) &= \mathrm{E}\left(e^{t(X_1 + \ldots + X_n)}\right) \\ &= M_X(t)^n \end{aligned}$$

This also holds true for $n = 0$, as both the left hand and right hand sides are 1. Thus we obtain

$$\mathrm{E}\left(e^{tS} \mid N\right) = M_X(t)^N$$

Using one of the basic properties of conditional expectation we then obtain

$$\begin{aligned} M_S(t) &= \mathrm{E}\left(e^{tS}\right) \\ &= \mathrm{E}\left(M_X(t)^N\right) \\ &= \mathrm{E}\left(e^{\log(M_X(t)))N}\right) \\ &= M_N\left(\log\left(M_X(t)\right)\right) \end{aligned}$$

Thus we get the mgf of S in terms of the mgf of N and the natural log of M_X. Note we may need to be careful of the appropriate domain of M_S. However this domain contains $(-\delta, \delta)$ for some $\delta > 0$.

End of Example

Another use of MGF's is to find moments. This is not always easier than finding the moments directly, but sometimes it is easier.

Consider the Gamma distribution. It has mgf

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$$

We obtain by differentiation

$$\begin{aligned}
\frac{dM(t)}{dt} &= \lambda^{\alpha}\alpha\,(\lambda - t)^{-(\alpha+1)} \\
\frac{d^2M(t)}{dt^2} &= \lambda^{\alpha}\alpha\,(\alpha+1)\,(\lambda - t)^{-(\alpha+2)} \\
\frac{d^3M(t)}{dt^3} &= \lambda^{\alpha}\alpha\,(\alpha+1)\,(\alpha+2)\,(\lambda - t)^{-(\alpha+3)} \\
\frac{d^4M(t)}{dt^4} &= \lambda^{\alpha}\alpha\,(\alpha+1)\,(\alpha+2)\,(\alpha+3)\,(\lambda - t)^{-(\alpha+4)}
\end{aligned}$$

From this we have

$$\begin{aligned}
\mathrm{E}(X) = M^{(1)}(0) &= \frac{\alpha}{\lambda} \\
\mathrm{E}(X^2) = M^{(2)}(0) &= \frac{\alpha\,(1+\alpha)}{\lambda^2} \\
\mathrm{E}(X^3) = M^{(3)}(0) &= \frac{\alpha\,(\alpha+1)\,(\alpha+2)}{\lambda^3} \\
\mathrm{E}(X^4) = M^{(4)}(0) &= \frac{\alpha\,(\alpha+1)\,(\alpha+2)\,(\alpha+3)}{\lambda^4}
\end{aligned}$$

Using a Taylor expansion about $t = 0$ of order 4 we obtain

$$\begin{aligned}
M(t) &= 1 + \frac{\alpha\,t}{\lambda} + \frac{\alpha\,(1+\alpha)\,t^2}{2\,\lambda^2} + \frac{\alpha\,\left(2 + 3\,\alpha + \alpha^2\right)\,t^3}{6\,\lambda^3} \\
&\quad + \frac{\alpha\,\left(6 + 11\,\alpha + 6\,\alpha^2 + \alpha^3\right)\,t^4}{24\,\lambda^4} + \mathrm{O}\,(t)^5
\end{aligned}$$

From this we also obtain

$$\begin{aligned}
\mathrm{E}(X) = M^{(1)}(0) &= \frac{\alpha}{\lambda} \\
\mathrm{E}(X^2) = M^{(2)}(0) &= \frac{\alpha\,(1+\alpha)}{\lambda^2} \\
\mathrm{E}(X^3) = M^{(3)}(0) &= \frac{\alpha\,\left(2 + 3\,\alpha + \alpha^2\right)}{\lambda^3} \\
\mathrm{E}(X^4) = M^{(4)}(0) &= \frac{\alpha\,\left(6 + 11\,\alpha + 6\,\alpha^2 + \alpha^3\right)}{\lambda^4}
\end{aligned}$$

We can obtain the MGF for a centred Gamma r.v. as

$$M_C(t) = e^{-t\alpha/\lambda}M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} e^{-\frac{t\,\alpha}{\lambda}}$$

We then obtain

$$M_C^{(1)}(t) = \frac{t\,\alpha \left(\frac{\lambda}{-t+\lambda}\right)^{\alpha}}{e^{\frac{t\,\alpha}{\lambda}}\,\left(-\left(t\,\lambda\right)+\lambda^2\right)}$$
$$= \frac{t\alpha}{\lambda}\frac{\left(\frac{\lambda}{-t+\lambda}\right)^{\alpha}}{e^{\frac{t\,\alpha}{\lambda}}\,\left(-t+\lambda\right)}$$

We obtain the skewness γ and kurtosis κ

$$\gamma = \frac{2}{\sqrt{\alpha}}$$
$$\kappa = 3+\frac{6}{\alpha}$$

Another interesting application of mgf's is to normalized sums of iid r.v.'s. This is a calculation that is needed in our study of the Central Limit Theorem.

Suppose that X_i, $i = 1, 2, \ldots$ are iid r.v.'s with an MGF, say M, and with $\mathrm{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = 1$. Consider the random variable

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \ .$$

Notice that $\mathrm{E}(Z_n) = 0$ and $\mathrm{Var}(Z_n) = 1$. What is the mgf of Z_n?

$$\begin{aligned} M_n(t) &= \mathrm{E}\left(e^{tZ_n}\right) \\ &= \mathrm{E}\left(e^{\frac{t}{\sqrt{n}} \sum_{i=1}^{n} X_i}\right) \\ &= \prod_{i=1}^{n} \mathrm{E}\left(e^{\frac{t}{\sqrt{n}} X_i}\right) \\ &= M\left(\frac{t}{\sqrt{n}}\right)^n \end{aligned}$$

Next we find find $\lim_{n\to\infty} M_n(t)$.

This limit can be found by using Taylor's Theorem with remainder. In the text, this method of mathematical argument is used, but without the careful study of the remainder in our approximation. For the purpose of this course that calculation is sufficient. However we can compute this limit is a valid fashion by taking advantage of L'Hôpital's Rule, a technique studied in a first calculus course.

Notice that we can take log's and study the limit of $\log(M_n(t))$. If this limit exists then $M_n(t)$ converges to the exponential of this limit. Next to study the limit we have

$$\begin{aligned} \lim_{n\to\infty} \log(M_n(t)) &= \lim_{n\to\infty} n \log\left(M\left(\frac{t}{\sqrt{n}}\right)\right) \\ &= \lim_{n\to\infty} \frac{1}{\frac{1}{n}} \log\left(M\left(\frac{t}{\sqrt{n}}\right)\right) \\ &= \lim_{\theta\to 0} \frac{\log\left(M\left(t\theta\right)\right)}{\theta^2} \end{aligned}$$

The later limit simply replaces $1/\sqrt{n}$ with θ in two places. It is valid in the sense that if the later limit exists, then the previous limit in terms of $n \to \infty$ also exists and is this same value. This method of embedding our limit into one that can make use of calculus allows us to obtain the limit using L'Hopital's Rule.

Using L'Hoptial's Rule twice we obtain

$$\begin{aligned} \lim_{\theta\to 0} \frac{\log\left(M(t\theta)\right)}{\theta^2} &= \lim_{\theta\to 0} \frac{M'(t\theta)t}{M(t\theta)2\theta} \\ &= \lim_{\theta\to 0} \frac{t}{2M(t\theta)} \lim_{\theta\to 0} \frac{M'\left(t\theta\right)}{\theta} \\ &= \frac{t}{2} \lim_{\theta\to 0} M''\left(t\theta\right) t \\ &= \frac{t^2}{2} \lim_{\theta\to 0} M''\left(t\theta\right) \end{aligned}$$

$$= \frac{t^2}{2} M''(0)$$
$$= \frac{t^2}{2} .$$

Thus $M_n(t) \to e^{\frac{t^2}{2}}$ as $n \to \infty$. This limit is the mgf of the $N(0,1)$ distribution. This result will be useful in understanding the Central Limit Theorem later in this course. The limit can also be obtained by an application of Taylor's formula.

We next find the limit

$$\lim_{n\to\infty} n \log \left(M \left(\frac{t}{\sqrt{n}} \right) \right)$$

using Taylor's approximtation. Since we evaluate M in the neighbourhood of 0 (since $\frac{t}{\sqrt{n}}$ is near 0 for large n) we use a Taylor's approximation

$$M(\theta) \approx M(0) + M'(0)\theta + \frac{1}{2} M''(0)\theta^2$$

We ignore the remainder term, but it can be made quite precise using Taylor's theorem with remainder. Thus we obtain

$$M(t/\sqrt{n}) \approx 1 + \frac{1}{2n} t^2$$

since $M(0) = 1, M'(0) = 0, M''(0) = 1$. Thus, using either L'Hopital's Rule in a manner similar to the above, or a Taylor's series for $\log(1+x)$ about $x = 0$, we obtain

$$\begin{aligned}
&\lim_{n\to\infty} n \log \left(M \left(\frac{t}{\sqrt{n}} \right) \right) \\
&= \lim_{n\to\infty} n \log \left(1 + \frac{1}{2n} t^2 \right) \\
&= \lim_{n\to\infty} n \frac{t^2}{2n} \\
&= \frac{t^2}{2} .
\end{aligned}$$