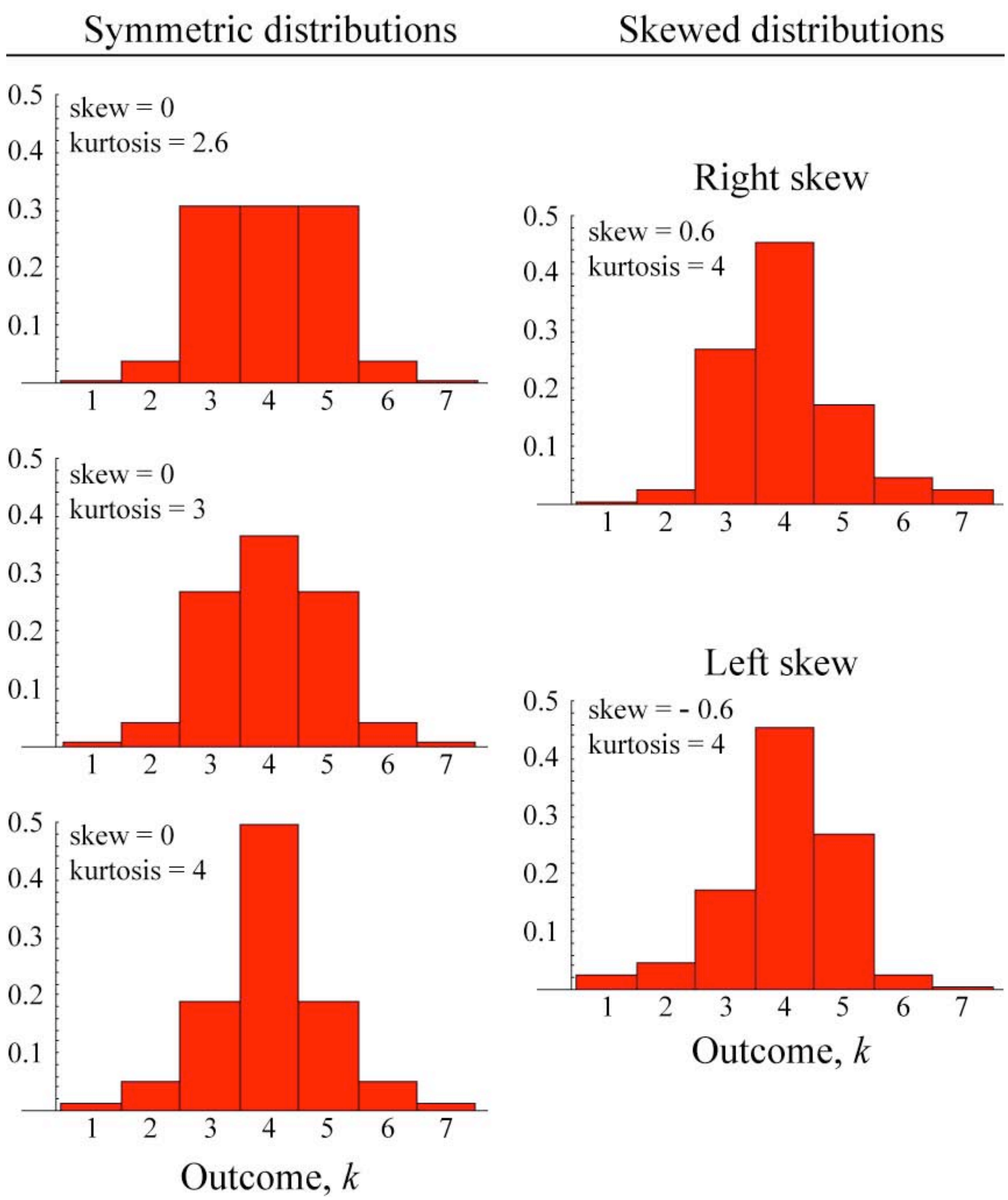
Appendix 5: Moment Generating Functions

Moment generating functions are used to calculate the mean, variance, and higher moments of a probability distribution. By definition, the j^{th} "*moment*" of a distribution is equal to the expected value of X^j. Thus, the first moment of a distribution is its mean, $E[X] = \mu$, and the second moment of a distribution, $E[X^2]$, is related to its variance (see equation P3.2 in Primer 3). Similarly, the j^{th} "*central moment*" of a distribution is defined as the expected value of $(X - \mu)^j$, which subtracts off the distribution's center of gravity (its mean) before calculating the expectation. The first central moment of a distribution is always zero, $E[X - \mu] = 0$, and the second central moment, $E[(X - \mu)^2]$, is the variance of the distribution.

While we have focused in the text on the first two moments, it is sometimes important to know the higher moments of a distribution. In particular, the third central moment, $E[(X - \mu)^3]$, provides a useful measure of the asymmetry of a distribution. This asymmetry is often measured by the "*skewness*", $E[(X - \mu)^3]/\sigma^3$, where σ is the standard deviation of the distribution (see Figure A5.1). Furthermore, the fourth central moment, $E[(X - \mu)^4]$, provides a useful measure of the peakedness of a distribution. This peakedness is often measured by the "*kurtosis*", $E[(X - \mu)^4]/\sigma^4$ (see Figure A5.1).

Figure A5.1: Skew and kurtosis of a distribution. A probability distribution with seven possible outcomes (k = 1…7) is used to illustrate skew and kurtosis. In each case, the mean was held at $\mu = 4$ and the variance at $\sigma^2 = 1$. The distributions on the left are symmetric (no skew), but the amount of kurtosis, $E[(X - \mu)^4]/\sigma^4$, increases from top to bottom. Distributions with a high kurtosis are more "peaked" and have "fatter" tails (i.e., a higher probability of $k = 1$ or $k = 7$). The distributions on the right have skew, $E[(X - \mu)^3]/\sigma^3$. The top distribution is "right-skewed" (skew > 0), with a fatter tail on the right. The bottom distribution is "left-skewed" (skew < 0), with a fatter tail on the left.



For any discrete distribution of interest, we could calculate these higher moments by summing x_i^j over the probability distribution for *X*:

$$E\left[X^j\right] = \sum_{x_i} x_i^j \; P\left(X = x_i\right) \; . \qquad \text{(A5.1)}$$

Calculating moment after moment can be tedious, but fortunately there is a simpler way to calculate all of the moments of a distribution. To so this we need to introduce *moment generating functions*. The moment generating function of a distribution is defined as:

$$MGF\left[z\right] = \sum_{x_i} e^{z\,x_i} \; P\left(X = x_i\right) \; . \qquad \text{(A5.2)}$$

Equation (A5.2) sums over all values that the random variable can take (x_i) and involves a newly introduced "dummy" variable, *z*. For some distributions, the sum in (A5.2) cannot be evaluated, in which case there is no point in using moment generating functions. But for many distributions, the sum in (A5.2) can be evaluated. For example, the moment generating function for the binomial distribution is known to equal $MGF\left[z\right] = \left(1 - p + e^z p\right)^n$, and Table P3.2 provides the moment generating functions for most of the discrete distributions introduced in Primer 3.

How do we use moment generating functions? Consider taking the derivative of (A5.2) with respect to the dummy variable *z*:

$$\frac{\mathrm{d}\left(MGF\left[z\right]\right)}{\mathrm{d}z} = \sum_{x_i} x_i \; e^{z\,x_i} \; P\left(X = x_i\right) \; . \qquad \text{(A5.3)}$$

If you then set the dummy variable, *z*, to zero, we regain the formula for the mean (Definition P3.2):

$$\left.\frac{\mathrm{d}\left(MGF\left[z\right]\right)}{\mathrm{d}z}\right|_{z=0} = \sum_{x_i} x_i \; P\left(X = x_i\right) \quad = E\left[X\right] \; . \qquad \text{(A5.4)}$$

Now consider taking the j^{th} derivative of (A5.2) with respect to the dummy variable z and setting z to zero, doing so gives the j^{th} moment of the distribution:

$$\left.\frac{\mathrm{d}^j\left(MGF[z]\right)}{\mathrm{d}z^j}\right|_{z=0} = \sum_{x_i} x_i^j \; P\left(X = x_i\right) \quad = E\left[X^j\right]. \tag{A5.5}$$

Furthermore, if we take the 0^{th} derivative, we get $MGF[0] = \sum_{x_i} P\left(X = x_i\right) = 1$, which reflects the fact that the sum over a probability distribution must equal one.

To see how amazing these results are, let's work with the moment generating function of the binomial distribution, $MGF[z] = \left(1 - p + e^z p\right)^n$. The derivative of this with respect to z is $n\, e^z p \left(1 - p + e^z p\right)^{n-1}$, which reduces to $n\ p$ when we set z to zero. Similarly, if we were interested in $E[X^3]$, we could calculate the third derivative with respect to z, set z to zero and find $E\left[X^3\right] = n\ p\left(1 - 3p + 3n\,p + 2p^2 - 3n\,p^2 + n^2p^2\right)$. As tedious as this is, it is much easier than having to figure out the sum in (A5.1).

Moment generating functions are extremely useful for showing how different distributions are related to one another. For example, we said in Primer 3 that the binomial distribution converges upon a Poisson distribution when p is small and n is large. In this case, the mean and variance of the binomial distribution are very nearly $\mu = n\,p$. This shows that the first two moments converge, but what about the higher moments? These also converge, as we can show by proving that the $MGF[z]$ of the binomial converges upon that of the Poisson distribution. The first step in the proof is to use the mean of the binomial, $\mu = n\,p$, to replace p with μ/n in the moment generating function of the binomial:

$$MGF[z] = \left(1 - \frac{\mu}{n} + e^z \frac{\mu}{n}\right)^n$$

In Box 7.4, we mentioned a remarkable relationship involving *e*, that $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$. We can use this relationship to see how the *MGF*[*z*] of the binomial changes as *n* gets large. By equating *x* to $\mu\left(e^z-1\right)$, the *MGF*[*z*] of the binomial converges upon $\lim_{n\to\infty}\left(1+\frac{\mu\left(e^z-1\right)}{n}\right)^n = e^{\mu\left(e^z-1\right)}$. But $e^{\mu\left(e^z-1\right)}$ is the moment generating function for the Poisson distribution. This proof, which treats μ as a constant, demonstrates that all of the moments of the binomial converge upon the moments of the Poisson as *n* gets large if we hold the mean constant at μ.

Often, we are not so interested in the j^{th} moment, $E\left[X^j\right]$, but rather in a related quantity, $E\left[\left(X-\mu\right)^j\right]$, known as the j^{th} *central moment*. For example, the variance of a distribution is the second central moment (Definition P3.3). To calculate central moments, we can just multiply the moment generating function by $e^{-z\,\mu}$. This gives us the *central moment generating function*, which can be written as

$$CMGF[z] = \sum_{x_i} e^{z\left(x_i-\mu\right)}\, P\left(X=x_i\right). \tag{A5.6}$$

Taking the j^{th} derivative of (A5.6) with respect to the dummy variable *z* and setting *z* to zero, gives the j^{th} central moment of the distribution:

$$\left.\frac{\mathrm{d}^j\left(CMGF[z]\right)}{\mathrm{d}z^j}\right|_{z=0} = \sum_{x_i}\left(x_i-\mu\right)^j\, P\left(X=x_i\right) \quad = E\left[\left(X-\mu\right)^j\right]. \tag{A5.7}$$

For example, multiplying the moment generating function for the binomial distribution by $e^{-z\,\mu}$ where μ is $n\,p$, we get the central moment generating function $CMGF[z] = e^{-z\,n\,p}\left(1-p+e^z p\right)^n$. Taking the first derivative of this with respect to *z* gives $e^{-z\,n\,p} n\, e^z p\left(1-p+e^z p\right)^{n-1} - \mu\, e^{-z\,n\,p}\left(1-p+e^z p\right)^n$, which reduces to 0 when we set *z* to zero and μ to $n\,p$. Of greater interest, we can apply this procedure to get the third central moment for the binomial distribution from the third derivative of $CMGF[z]$. Doing so, we find that $E\left[(X-\mu)^3\right] = n\ p\,(1-p)\,(1-2p)$. This tells us that the binomial distribution is skewed unless $p = 1/2$.

The same principles apply to continuous distributions, but the above sums become integrals. Specifically, the moment generating function for a continuous distribution is defined as:

$$MGF[z] = \int_a^b e^{z\,x} f(x)\,\mathrm{d}x \ . \tag{A5.8}$$

Again, the integral in (A5.8) is evaluated over the entire range of values that the random variable can take (a, b). The central moment generating function is obtained by multiplying the moment generating function by $e^{-z\,\mu}$, if $MGF[z]$ is already known, or by direct calculation:

$$CMGF[z] = \int_a^b e^{z\,(x-\mu)} f(x)\,\mathrm{d}x \, . \tag{A5.9}$$

Table P3.3 provides the moment generating functions for most of the continuous distributions discussed in Primer 3. As an example, the moment generating function for the normal distribution is defined as $MGF[z] = \int_{-\infty}^{\infty} e^{z\,x} \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}\,\mathrm{d}x$, which equals $MGF[z] = e^{z\,\mu + z^2\sigma^2/2}$. Because $CMGF[z] = e^{-z\,\mu}\,MGF[z]$, the central moment generating function for the normal distribution is even simpler: $CMGF[z] = e^{z^2\sigma^2/2}$. Using the central moment generating function, it becomes easy to calculate the first four central moments for the normal distribution:

$$E[(X-\mu)] = \left.\frac{\mathrm{d}e^{z^2\sigma^2/2}}{\mathrm{d}z}\right|_{z=0} = 0 \ . \tag{A5.10a}$$

$$E\left[(X-\mu)^2\right] = \left.\frac{\mathrm{d}^2 e^{z^2\sigma^2/2}}{\mathrm{d}z^2}\right|_{z=0} = \sigma^2 \ . \tag{A5.10b}$$

$$E\left[(X-\mu)^3\right] = \left.\frac{\mathrm{d}^3 e^{z^2\sigma^2/2}}{\mathrm{d}z^3}\right|_{z=0} = 0 \ . \tag{A5.10c}$$

$$E\left[(X-\mu)^4\right] = \left.\frac{\mathrm{d}^4 e^{z^2\sigma^2/2}}{\mathrm{d}z^4}\right|_{z=0} = 3\sigma^4 \ . \tag{A5.10d}$$

We can then conclude that the normal distribution has no skew, while it has a kurtosis equal to $\frac{E\left[(X-\mu)^4\right]}{\sigma^4} = 3$. Only derivatives are needed in these calculations, allowing us to avoid any further integrals once the moment generating function has been calculated.

Before leaving the topic, one weakness and one strength of moment generating functions deserve mention. A weakness of moment generating functions is that they do not always exist, because e^{zx} can grow so fast when multiplied by $f(x)$ that the sum in (A5.2) (or the integral in (A5.8)) is infinite. This problem can be circumvented using the *characteristic function* of a distribution, which multiplies $f(x)$ by e^{izx} instead of e^{zx}. The characteristic function always converges and moments can be determined in a similar fashion, but the introduction of complex numbers is unnecessary for most problems. One powerful advantage of moment generating functions is that they can be combined and manipulated quickly to demonstrate important facts about probability distributions:

Rule A5.1: Properties of moment generating functions

(a) If the random variable X has the moment generating function $MGF[z]$, then $Y = a\,X + b$ has a moment generating function $e^{z\,b}\,MGF[a\,z]$.

(b) The moment generating function for a sum of independent random variables is the product of each moment generating function.

Rule A5.1 applies equally to central moment generating functions.

Rule A5.1a can be used to calculate the effect of scaling and shifting a probability distribution on moments like the mean and the variance. For example, the binomial distribution is often used to describe how genetic drift affects the number of alleles, *A*, in a population of *N* alleles, when *p* is the frequency of allele *A* before drift (see Chapters 14 and 15). Because the *number* of *A* alleles after genetic drift (*X*) follows a binomial distribution with $CMGF[z] = e^{-zNp}\left(1-p+e^{z}p\right)^{N}$, Rule A5.1a tells us that the *frequency* of allele *A* after genetic drift ($Y = X/N$) follows a distribution with $CMGF[z] = e^{-zp}\left(1-p+e^{z/N}p\right)^{N}$. By differentiating $CMGF[z]$ with respect to *z* and setting *z* to zero, the first three central moments of the frequency of allele *A* after drift are 0, $\frac{p(1-p)}{N}$, and $\frac{p(1-p)(1-2p)}{N^2}$. Consequently, there is no expected change in the allele frequency (from the first moment), but the variance in allele frequency is proportional to $1/N$ (from the second moment) and is thus smaller in larger populations. Higher order moments, including the third central moment, are proportional to $1/N^2$, which is much smaller and can be neglected assuming that the population is reasonably large.

Rule A5.1b allows you to calculate moments for compound probability distributions, involving more than one random event. For example, if the trait of an individual reflects the sum effect of a genetic influence, summarized by a normal distribution with mean μ_1 and variance σ_1^2, and an environmental influence, summarized by a normal distribution with mean μ_2 and variance σ_2^2, the moment generating function for the trait will be the product of the moment generating functions for the genetic and environmental influences:

$$\begin{aligned} MGF[z] &= e^{z\,\mu_1 + z^2\sigma_1^2/2} e^{z\,\mu_2 + z^2\sigma_2^2/2} \\ &= e^{z(\mu_1+\mu_2) + z^2(\sigma_1^2+\sigma_2^2)/2} \end{aligned} \quad . \tag{A5.11}$$

Because (A5.11) is the moment generating function of a normal distribution with mean ($\mu_1 + \mu_2$) and variance ($\sigma_1^2 + \sigma_2^2$), the trait will remain normally distributed, even though it is influenced by multiple factors. This is a relatively painless way to prove that the sum of normally distributed random variables is itself normally distributed.

Exercise A5.1:

(a) Use equation (A5.8) to prove that the moment generating function of a uniform distribution is $MGF[z] = \frac{e^{\max z} - e^{\min z}}{(\max - \min)\, z}$. (b) Calculate the mean of the uniform distribution from $E[X] = \left.\frac{\mathrm{d}(MGF[z])}{\mathrm{d}z}\right|_{z=0}$. [Note: You will need to use L'Hôpital's Rule A2.30 to calculate the limit as z goes to zero.]

[Answer[1]]

Exercise A5.2:

(a) Use the moment generating function of the Poisson to show that the sum of two Poisson distribution with means μ_1 and μ_2 also follows a Poisson distribution. (b) If the rate of offspring production is 0.1 per year during the first five years of reproductive life and then becomes 0.2 per year during the next five years, what are the mean and variance of the sum total number of offspring born per parent over the ten years?

[Answer[2]]

References:

Stuart, A., and J. K. Ord. 1987. Kendall's advanced theory of statistics, Vol. 1. Distribution theory. Oxford University Press, Oxford.

---

[1] ANSWER:

(a) Using the probability density function for a uniform distribution, $f(x) = \frac{1}{\max - \min}$, equation (A5.8) becomes $MGF[z] = \int_{\max}^{\min} \frac{e^{z\,x}}{\max - \min}\,\mathrm{d}x = \left.\frac{e^{z\,x}}{(\max - \min)\,z}\right|_{\max}^{\min}$, which correctly evaluates to the moment generating function of the uniform distribution, $MGF[z] = \frac{e^{\max z} - e^{\min z}}{(\max - \min)\,z}$.

(b) For a uniform distribution, $E[X] = \left.\frac{\mathrm{d}(MGF[z])}{\mathrm{d}z}\right|_{z=0}$ equals

$\frac{1}{(\max - \min)} \left.\frac{\left(\max e^{\max z} - \min e^{\min z}\right) z - \left(e^{\max z} - e^{\min z}\right)}{z^2}\right|_{z=0}$. Here, we run into a problem, because both the numerator and the denominator approach zero as z goes to zero. Fortunately, we can use L'Hôpital's Rule (A2.30) to evaluate this function, allowing us to rewrite

$\frac{1}{(\max - \min)} \lim_{z \to 0} \frac{\left(\max e^{\max z} - \min e^{\min z}\right) z - \left(e^{\max z} - e^{\min z}\right)}{z^2}$ as

$\frac{1}{(\max - \min)} \lim_{z \to 0} \frac{\left(\max^2 e^{\max z} - \min^2 e^{\min z}\right) z}{2\,z}$. Canceling out the z in the numerator and denominator and taking the limit, we get the mean of the uniform distribution:

$\frac{\left(\max^2 - \min^2\right)}{2(\max - \min)} = \frac{(\max + \min)}{2}$. While it is much easier to calculate the mean for the uniform distribution directly (see Primer 3), this example illustrates what to do in cases where both the numerator and denominator equal zero in a moment generating function or its derivatives.

[2] ANSWER:

(a) According to rule A5.1b, the moment generating function for the sum of two Poisson distributions is given by $MGF[z] = e^{\mu_1 (e^z - 1)} e^{\mu_2 (e^z - 1)}$, which equals $MGF[z] = e^{(\mu_1 + \mu_2)(e^z - 1)}$. This moment generating function is the moment generating function of a Poisson distribution with mean $(\mu_1 + \mu_2)$.

(b) The number of offspring during the first five years is Poisson distributed with mean 0.5 (= 0.1 x 5), and the number of offspring during the next five years is Poisson distributed with mean 1.0 (= 0.2 x 5). The sum total number of offspring is thus Poisson distributed with mean 1.5 offspring. Because the variance of the Poisson is equal to the mean, the variance is also 1.5. In this example, the rate of events varied over time, but in a known fashion. In this case, the total number of events remained Poisson distributed with mean given by the expected number of events during the whole time interval, $\left(0.1\frac{5}{10} + 0.2\frac{5}{10}\right) \times 10 = 1.5$. If, however, the expected number of events was itself a random variable (e.g., was gamma distributed with mean 1.5), then the process would no longer be Poisson (see Stuart and Ord 1987, p. 182).