CLARK UNIVERSITY 1887

Moments and the moment generating function
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

There are various reasons for studying moments and the moment generating functions. One of them that the moment generating function can be used to prove the central limit theorem.

Moments, central moments, skewness, and kurtosis. The k^{th} *moment* of a random variable X is defined as $\mu_k = E(X^k)$. Thus, the mean is the first moment, $\mu = \mu_1$, and the variance can be found from the first and second moments, $\sigma^2 = \mu_2 - \mu_1^2$.

The k^{th} *central moment* is defined as $E((X-\mu)^k)$. Thus, the variance is the second central moment.

The higher moments have more obscure meanings as k grows.

A third central moment of the standardized random variable $X^* = (X-\mu)/\sigma$,

$$\beta_3 = E((X^*)^3) = \frac{E((X-\mu)^3)}{\sigma^3}$$

is called the *skewness* of X. A distribution that's symmetric about its mean has 0 skewness. (In fact all the odd central moments are 0 for a symmetric distribution.) But if it has a long tail to the right and a short one to the left, then it has a positive skewness, and a negative skewness in the opposite situation.

A fourth central moment of X^*,

$$\beta_4 = E((X^*)^4) = \frac{E((X-\mu)^4)}{\sigma^4}$$

is callled *kurtosis*. A fairly flat distribution with long tails has a high kurtosis, while a short tailed distribution has a low kurtosis. A bimodal distribution has a very high kurtosis. A normal distribution has a kurtosis of 3. (The word kurtosis was made up in the early 19th century from the Greek word for curvature.) Kurtosis is not a particularly important concept, but I mention it here for completeness.

It turns out that the whole distribution for X is determined by all the moments, that is different distributions can't have identical moments. That's what makes moments important.

The moment generating function. There is a clever way of organizing all the moments into one mathematical object, and that object is called the *moment generating function*. It's a function $m(t)$ of a new variable t defined by

$$m(t) = E(e^{tX}).$$

Since the exponential function e^t has the power series

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^k}{k!} + \cdots,$$

we can rewrite $m(t)$ as follows

$$\begin{aligned} m(t) &= E(e^{tX}) \\ &= E\left(1 + tX + \frac{(tX)^2}{2!} + \cdots + \frac{(tX)^k}{k!} + \cdots\right) \\ &= 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \cdots + \frac{t^kE(X^k)}{k!} + \cdots \\ &= 1 + t\mu_1 + \frac{t^2\mu_2}{2!} + \cdots + \frac{t^k\mu_k}{k!} + \cdots \\ &= 1 + \mu_1 t + \frac{\mu_2}{2!}t^2 + \cdots + \frac{\mu_k}{k!}t^k + \cdots \end{aligned}$$

Theorem. The k^{th} derivative of $m(t)$ evaluated at $t=0$ is the k^{th} moment μ_k of X.

In other words, the moment generating function generates the moments of X by differentiation.

The primary use of moment generating functions is to develop the theory of probability. For instance, the easiest way to prove the central limit theorem is to use moment generating functions.

For discrete distributions, we can also compute the moment generating function directly in terms of the probability mass function $f(x) = P(X{=}x)$ as

$$m(t) = E(e^{tX}) = \sum_x e^{tx} f(x).$$

For continuous distributions, the moment generating function can be expressed in terms of the probability density function f as

$$m(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx.$$

Properties of moment generating functions.
Translation. If $Y = X + a$, then

$$m_Y(t) = e^{at} m_X(t).$$

Proof: $m_Y(t) = E(e^{Yt}) = E(e^{(X+a)t}) = E(e^{Xt}e^{at}) = e^{at}E(e^{Xt}) = e^{at}m_X(t)$. Q.E.D.

Scaling. If $Y = bX$, then

$$m_Y(t) = m_X(bt).$$

Proof: $m_Y(t) = E(e^{Yt}) = E(e^{(bX)t}) = E(e^{X(bt)}) = m_X(bt)$. Q.E.D.

Standardizing. From the last two properties, if

$$X^* = \frac{X - \mu}{\sigma}$$

is the standardized random variable for X, then

$$m_{X^*}(t) = e^{-\mu t/\sigma} m_X(t/\sigma).$$

Proof: First translate by $-\mu$ to get

$$m_{X-\mu}(t) = e^{-\mu t} m_X(t).$$

Then scale that by a factor of $1/\sigma$ to get

$$\begin{aligned} m_{(X-\mu)/\sigma} &= m_{X-\mu}(t/\sigma) \\ &= e^{-\mu t/\sigma} m_X(t/\sigma) \end{aligned}$$

Q.E.D.

Convolution. If X and Y are independent variables, and $Z = X + Y$, then

$$m_Z(t) = m_X(t)\, m_Y(t).$$

Proof: $m_Z(t) = E(e^{Zt}) = E(e^{(X+Y)t}) = E(e^{Xt}e^{Yt})$. Now, since X and Y are independent, so are e^{Xt} and e^{Yt}. Therefore, $E(e^{Xt}e^{Yt}) = E(e^{Xt})E(e^{Yt}) = m_X(t)m_Y(t)$. Q.E.D.

Note that this property of convolution on moment generating functions implies that for a sample sum $S_n = X_1 + X_2 + \cdots + X_n$, the moment generating function is

$$m_{S_n}(t) = (m_X(t))^n.$$

We can couple that with the standardizing property to determine the moment generating function for the standardized sum

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Since the mean of S_n is $n\mu$ and its standard deviation is $\sigma\sqrt{n}$, so when it's standardized, we get

$$\begin{aligned} m_{S_n^*}(t) &= e^{-n\mu t/(\sigma\sqrt{n})} m_{S_n}(\tfrac{t}{\sigma\sqrt{n}}) \\ &= e^{-\sqrt{n}\mu t/\sigma} m_{S_n}(\tfrac{t}{\sigma\sqrt{n}}) \\ &= e^{-\sqrt{n}\mu t/\sigma} \left(m_X(\tfrac{t}{\sigma\sqrt{n}}) \right)^n \end{aligned}$$

We'll use this result when we prove the central limit theorem

Some examples. The same definitions for these apply to discrete distributions and continuous ones. That is, if X is any random variable, then its n^{th} moment is

$$\mu_n = E(X^n)$$

and its moment generating function is

$$m_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} \frac{\mu_k}{k!}\, t^k.$$

The only difference is that when you compute them, you use sums for discrete distributions but integrals for continuous ones. That's because expectation is defined in terms of sums or integrals in the two cases. Thus, in the continuous case

$$\mu_n = E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x)\, dx$$

and

$$m(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx.$$

We'll look at one example of a moment generating function for a discrete distribution and three for continuous distributions.

The moment generating function for a geometric distribution. Recall that when independent Bernoulli trials are repeated, each with probability p of success, the time X it takes to get the first success has a geometric distribution.

$$P(X = j) = q^{j-1}p, \qquad \text{for } j = 1, 2, \ldots.$$

Let's compute the generating function for the geometric distribution.

$$\begin{aligned} m(t) &= \sum_{j=1}^{\infty} e^{tj} q^{j-1} p \\ &= \frac{p}{q} \sum_{j=1}^{\infty} e^{tj} q^j \end{aligned}$$

The series $\displaystyle\sum_{j=1}^{\infty} e^{tj} q^j = \sum_{j=1}^{\infty} (e^t q)^j$ is a geometric series with sum $\dfrac{e^t q}{1 - e^t q}$. Therefore,

$$m(t) = \frac{p}{q} \frac{e^t q}{1 - e^t q} = \frac{pe^t}{1 - qe^t}.$$

From this generating function, we can find the moments. For instance, $\mu_1 = m'(0)$. The derivative of m is

$$m'(t) = \frac{pe^t}{(1 - qe^t)^2},$$

so $\mu_1 = m'(0) = \dfrac{p}{(1-q)^2} = \dfrac{1}{p}$. This agrees with what we already know, that the mean of the geometric distribution is $1/p$.

The moment generating function for a uniform distribution on $[0, 1]$**.** Let X be uniform on $[0, 1]$ so that the probability density function f_X has the value 1 on $[0, 1]$ and 0 outside this interval.

Let's first compute the moments.

$$\begin{aligned} \mu_n = E(X^n) &= \int_{-\infty}^{\infty} x^n f_X(x)\, dx \\ &= \int_0^1 x^n\, dx \\ &= \left. \frac{x^{n+1}}{n+1} \right|_0^1 = \frac{1}{n+1} \end{aligned}$$

Next, let's compute the moment generating function.

$$\begin{aligned} m(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx \\ &= \int_0^1 e^{tx}\, dx \\ &= \left. \frac{1}{t} e^{tx} \right|_0^1 = \frac{e^t - 1}{t} \end{aligned}$$

Note that the expression for $g(t)$ does not allow $t = 0$ since there is a t in the denominator. Still $g(0)$ can be evaluated by using power series or L'Hôpital's rule.

The moment generating function for an exponential distribution with parameter λ**.** Recall that when events occur in a Poisson process uniformly at random over time at a rate of λ events per unit time, then the random variable X giving the time to the first event has an exponential distribution. The density function for X is $f_X(x) = \lambda e^{-\lambda x}$, for $x \in [0, \infty)$.

Let's compute its moment generating function.

$$\begin{aligned} m(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\,dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x}\,dx \\ &= \lambda \frac{e^{(t-\lambda)x}}{t-\lambda}\bigg|_0^{\infty} \\ &= \left(\lim_{x\to\infty} \lambda \frac{e^{(t-\lambda)x}}{t-\lambda}\right) - \lambda \frac{e^0}{t-\lambda} \end{aligned}$$

Now if $t < \lambda$, then the limit in the last line is 0, so in that case

$$m(t) = \frac{\lambda}{\lambda - t}.$$

This is a minor, yet important point. The moment generating function doesn't have to be defined for all t. We only need it to be defined for t near 0 because we're only interested in its derivatives evaluated at 0.

The moment generating function for the standard normal distribution. Let Z be a random variable with a standard normal distribution. Its probability density function is

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

Its moments can be computed from the definition, but it takes repeated applications of integration by parts to compute

$$\mu_n = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, x^n e^{-x^2/2}\,dx.$$

We won't do that computation here, but it turns out that when n is odd, the integral is 0, so μ_n is 0 if n is odd. On the other hand when n is even, say $n = 2m$, then it turns out that

$$\mu_{2m} = \frac{(2m)!}{2^m m!}.$$

From these values of all the moments, we can compute the moment generating function.

$$\begin{aligned} m(t) &= \sum_{n=0}^{\infty} \frac{\mu_n}{n!}\, t^n \\ &= \sum_{m=0}^{\infty} \frac{\mu_{2m}}{(2m)!}\, t^{2m} \\ &= \sum_{m=0}^{\infty} \frac{(2m)!}{2^m m!}\, \frac{1}{(2m)!}\, t^{2m} \\ &= \sum_{m=0}^{\infty} \frac{1}{2^m m!}\, t^{2m} \\ &= e^{t^2/2} \end{aligned}$$

Thus, the moment generating function for the standard normal distribution Z is

$$m_Z(t) = e^{t^2/2}.$$

More generally, if $X = \sigma Z + \mu$ is a normal distribution with mean μ and variance σ^2, then the moment generating function is

$$g_X(t) = \exp(\mu t + \sigma^2 t^2/2).$$

Math 217 Home Page at
`http://math.clarku.edu/~djoyce/ma217/`