THE CENTRAL LIMIT THEOREM AND MOMENT GENERATING FUNCTIONS

Maurice J. Dupré
Department of Mathematics
New Orleans, LA 70118
email: mdupre@tulane.edu
November 2010

The **Central Limit Theorem** says that in the limit as the sample size becomes infinite, with independent random sampling, the distribution of the sample total and the sample mean both become normal. To state this precisely requires a fair amount of sophisticated mathematics which we will not bother with. We do know that for independent random sampling of the random variable X we generate a sequence of unknowns $X_1, X_2, X_3, ..., X_n, ...$ which we can think of as potentially infinite, and as these are all observations of X, they all have the same distribution as X. In particular, they all have the same mean, μ, and standard deviation, σ, as X, that is to say,

$$\mu_{X_k} = \mu_X, \ \ k = 1, 2, 3, ..., n, ...$$

and

$$\sigma_{X_k} = \sigma_X, \ \ k = 1, 2, 3, ..., n, ...$$

But, in addition, we assume that X_i is independent of X_j if $i \neq j$. When we take a sample of size n, we usually are interested in the sample mean as an estimate of the the true mean, and to form the sample mean we begin by forming the sample total T_n given by

$$T_n = \sum_{k=0}^{n} X_k,$$

and then divide by the sample size to get the sample mean $\bar{X}_n$, so

$$\bar{X}_n = \frac{1}{n} T_n.$$

Now from properties of expectation, we know that for any sampling method, we have

$$E(T_n) = n\mu_X$$

and therefore

$$E(\bar{X}_n) = \mu_X.$$

But for independent random sampling, we have $Var(T_n)$, the variance for T_n, given by

$$Var(T_n) = n \cdot \sigma_X^2,$$

and therefore $SD(T_n)$, its standard deviation, is given by

$$SD(T_n) = (\sqrt{n}) \cdot \sigma_X.$$

Consequently, for independent random sampling we have

$$SD(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}},$$

and therefore

$$Var(\bar{X}_n) = \frac{\sigma_X^2}{n}.$$

Now, the fact that

$$SD(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}}$$

means that, as $n \to \infty$, the standard deviation of $\bar{X}_n$ goes to zero indicating that in the limit we must have just a constant. This certainly does not seem conducive to finding the distribution of $\bar{X}_n$ as $n \to \infty$. The way around this is to standardize. Recall that the standardization of any random variable or unknown X, denoted Z_X, is given by

$$Z_X = \frac{X - \mu_X}{\sigma_X},$$

so in terms of Z_X we recover X as

$$X = \mu_X + \sigma_X \cdot Z_X.$$

In particular,

$$Z_{T_n} = \frac{T_n - n\mu_X}{\sqrt{n} \cdot \sigma_X},$$

so when we multiply numerator and denominator by $1/n$ we find

$$Z_{T_n} = \frac{\bar{X}_n - \mu_X}{(\frac{\sigma_X}{\sqrt{n}})} = Z_{\bar{X}_n}.$$

The more precise statement of the Central Limit Theorem says that as $n \to \infty$, the standardization $Z_{T_n} = Z_{\bar{X}_n}$, converges in some sense to a standard normal random variable.

To make this plausible, we pass to the moment generating function. Recall that for any random variable X, its moment generating function is m_X, where

$$m_X(t) = E(\exp(tX)) = \int_{-\infty}^{\infty} \exp(xt) dF_X(x).$$

Of course, F_X is the cumulative distribution function of X, so

$$F_X(x) = P(X \leq x), \; x \in \mathbb{R}.$$

Moreover, it is the case that if $m_X = m_Y$, for unknowns X and Y, then $F_X = F_Y$, that is if two unknowns have the same moment generating function then they have the same distribution. And, if X and Y are independent, then $\exp(tX)$ and $\exp(tY)$ are uncorrelated and therefore

$$E(e^{tX} \cdot e^{tY}) = [E(e^{tX})] \cdot [E(e^{tY})],$$

so

$$m_{[X+Y]}(t) = E(e^{t(X+Y)}) = [E(e^{tX})] \cdot [E(e^{tY})] = m_X(t) \cdot m_Y(t),$$

or

$$m_{[X+Y]} = m_X \cdot m_Y,$$

if X and Y are independent. Thus the moment generating function of any sum of independent unknowns is simply the product of the moment generating functions of all the summands. If we denote differentiation with respect to t with prime as in

$$\frac{dg}{dt}(t) = g'(t),$$

then

$$m'_X(t) = E(Xe^{tX}),$$

$$m''_X(t) = E(X^2e^{tX}),$$

and so on. It is convenient to denote the n^{th} derivative of g by $g^{(n)}$ when dealing with higher derivatives, so we see that

$$m_X^{(n)}(t) = E(X^n e^{tX}),\ n = 0, 1, 2, 3, ...,$$

and putting $t = 0$ into these derivatives gives

$$m_X^{(n)} = E(X^n),\ n = 0, 1, 2, 3, ...$$

showing where the moment generating function gets its name. The numbers $E(X^n)$ are called the moments of X, and we see that the moment generating function generates them all via differentiation and evaluation at zero. Notice that $X^0 = 1$, and $m_X(0) = E(1) = 1$. Now it can be shown that the moment generating function equals its Taylor expansion about $t = 0$ for t close enough to zero, so there is $\epsilon > 0$ with

$$m_X(t) = \sum_{n=0}^{\infty} m_X^{(n)}(0) \cdot \frac{t^n}{n!},\ |t| < \epsilon.$$

In the case where X is a continuous random variable, it has pdf (probability density function)

$$f_X = \frac{d}{dx} F_X,$$

so $dF_X(x) = f_X(x)dx$ and

$$m_X(t) = E(\exp(tX)) = \int_{-\infty}^{\infty} \exp(xt) f_X(x) dx.$$

If Z is a standard normal random variable, then its pdf is given by

$$f_Z(z) = \frac{\exp(-z^2/2)}{\sqrt{2\pi}},\ z \in \mathbb{R}.$$

A calculation shows that if Z is any standard normal random variable,

$$m_Z(t) = \exp\left(\frac{t^2}{2}\right).$$

This means that we can make the Central Limit Theorem plausible if we can show that the moment generating function of Z_{T_n} converges to m_Z as $n \to \infty$.

Now, for moment generating functions, we see that

$$m_{X\pm c}(t) = E(e^{(X\pm c)t} = E(e^{\pm ct} \cdot e^{tX}) = e^{\pm ct} m_X(t),$$

so

$$m_{X\pm c} = e^{ct} \cdot m_X.$$

Also, as regards rescaling,

$$m_{cX}(t) = E(e^{t(cX)}) = E(e^{(ct)X}) = m_X(ct).$$

Now to see the Central Limit Theorem, let us begin with a standard random variable but otherwise having any distribution what so ever, and call it X. Let Z denote a standard normal random variable. Let us consider an infinite sequence of independent unknowns $X_1, X_2, X_3, ..., X_n, ...$ all having the same distribution as X, so all are standard and thus

$$E(X_k) = 0, \ k = 1, 2, 3, ...$$

and

$$E(X_k^2) = 1, \ k = 1, 2, 3, ...$$

and set $T = T_n$ and $\bar{X}_n$. Thus we have

$$m_{X_k} = m_X, \ k = 1, 2, 3, ...$$

and

$$m_T = m_{X_1} \cdot m_{X_2} \cdot m_{X_3} \cdots m_{X_n},$$

which together gives

$$m_T = (m_X)^n.$$

But now, to get the proper handle on the distribution of T, we already remarked we need to standardize T. But from our formulas above for mean and standard deviation of T_n and $\bar{X}_n$ we have, as here $E(X) = 0$ and $\sigma_X = 1$,

$$\sigma_T = \sqrt{n}$$

and therefore

$$Z_T = \frac{T}{\sqrt{n}}.$$

This means that

$$m_{Z_{T_n}}(t) = m_{T_n}\left(\frac{t}{\sqrt{n}}\right) = \left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

To get the limit distribution we need to take the limit as $n \to \infty$ of the expression on the right side above. Of course, that is obviously a problem since the $\sqrt{n}$ in the denominator would seem to be wiping out the dependence on t, but in a sense, that is really just what we need. There are certainly technical difficulties with taking the limit as $n \to \infty$ here, but we are only looking to make a plausibiity argument here, so let us assume that we can take the limit as $n \to \infty$. In fact, let us assume the limit function h is differentiable and that differentiation can commute with the limit process here. Thus, lets assume

$$h(t) = \lim_{n \to \infty} \left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

and

$$\frac{d}{dt}h(t) = \lim_{n \to \infty} \frac{d}{dt}\left(\left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^n\right).$$

Next, we can calculate the derivative on the right side using the Chain Rule getting

$$\frac{d}{dt}\left(\left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^n\right) = n \cdot \left(\left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^{n-1}\right) \cdot m_X'\left(\frac{t}{\sqrt{n}}\right) \cdot \frac{1}{\sqrt{n}}$$

$$= \left(\left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^{n-1}\right) \cdot m_X'\left(\frac{t}{\sqrt{n}}\right) \cdot \sqrt{n}.$$

That is, if we set

$$h_n = \left[m_X\left(\frac{t}{\sqrt{n}}\right)\right]^n, \; n = 1, 2, 3, ...,$$

then

$$h(t) = \lim_{n\to\infty} h_n(t),$$

and

$$\frac{d}{dt}h(t) = \lim_{n\to\infty}\left[h_{n-1}(t) \cdot m_X'\left(\frac{t}{\sqrt{n}}\right) \cdot \sqrt{n}\right].$$

Now certainly,

$$\lim_{n\to\infty} h_{n-1}(t) = \lim_{n\to\infty} h_n(t) = h(t),$$

so the limit calculation boils down to the calculation of the limit

$$L = \lim_{n\to\infty} m_X'\left(\frac{t}{\sqrt{n}}\right) \cdot \sqrt{n}.$$

If L exists, then

$$\frac{d}{dt}h(t) = L \cdot h(t).$$

Let us put

$$\delta = \frac{1}{\sqrt{n}},$$

so then we instead need to calculate L where

$$L = \lim_{\delta\to 0} \frac{m_X'(\delta \cdot t)}{\delta}.$$

Now, putting $\delta = 0$ in the numerator gives $m_X'(0) = E(X) = 0$, so if we define the function g by

$$g(\delta) = m_X'(\delta \cdot t),$$

then $g(0) = 0$, and its derivative is, from the limit definition of derivative,

$$g'(0) = \lim_{\delta\to 0} \frac{g(0+\delta) - g(0)}{\delta} = \lim_{\delta\to 0} \frac{g(\delta)}{\delta} = \lim_{\delta\to 0} \frac{m_X'(\delta \cdot t)}{\delta} = L.$$

We therefore have

$$L = g'(0).$$

But since

$$g(\delta) = m_X'(\delta \cdot t),$$

we may instead use the Chain Rule, which results in

$$g'(\delta) = \frac{d}{d\delta}[m_X'(\delta \cdot t)] = t \cdot m_X''(\delta \cdot t),$$

and therefore

$$L = g'(0) = t \cdot m''_X(0).$$

Next, recall that the reason m_X is called the moment generating function is that the n^{th} derivative at zero is the n^{th} moment, $E(X^n)$. Thus,

$$m''_X(0) = E(X^2) = 1,$$

since X is assumed to be standard. The end result is that we have simply

$$L = t.$$

But recall now that we have $h' = L \cdot h$, so this means we now have simply

$$h'(t) = th(t).$$

If we assume that h is the moment generating function for some unknown, then it is equal to its power series in a neighborhood of $t = 0$, so we can find h from its power series. We must calculate all derivatives of h at $t = 0$. Notice that as $h'(t) = th(t)$, if h is differentiable, then it must have a second derivative and by the Product Rule for differentiation,

$$h''(t) = h(t) + th'(t) = h(t) + t^2h(t) = [1 + t^2]h(t),$$

and then as h is differentiable, this last equation shows that h has a third derivative. In fact, the equation $h' = th$ shows that if h has an n^{th} derivative, then it must have an $(n+1)^{th}$ derivative, so we can see that the assumption that h equals its power series expansion about $t = 0$ is not completely out of bounds. Notice that we can determine all the derivatives of h by repeatedly differentiating $h' = th$, and for each n we find a polynomial $p_n(t)$ so that

$$h^{(n)}t) = p_n(t) \cdot h(t).$$

We have

$$h_n(t) = \left[m_X(\left(\frac{t}{\sqrt{t}} \right) \right]^n,$$

so

$$h_n(0) = [m_X(0)]^n = 1^n = 1,$$

and therefore

$$h(0) = \lim_{n \to \infty} h_n(0) = 1,$$

so

$$h^{(n)}(0) = p_n(0)h(0) = p_n(0), \ n = 0, 1, 2, 3, ...$$

Notice that these polynomials come merely from the fact that $h' = th$. Any function equal to a power series which satisfies this equation must then have the exact same power series and so must equal h. But, if we look at m_Z, the moment generating function of the standard normal random variable Z, then

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

so by the Chain Rule for differentiation we have

$$m'_Z(t) = e^{t^2/2} \cdot \frac{2t}{2} = te^{t^2/2} = tm_Z(t),$$

the exact same equation, and $m_Z(0) = 1$, as well. As a consequence, it must be that also

$$m_Z^{(n)}(0) = p_n(0) = h^{(n)}(0), \ n = 0, 1, 2, 3, ...$$

Thus, both m_Z and h have the same power series expression and are therefore the exact same function, so finally, we have the Central Limit Theorem:

$$m_Z = \lim_{n\to\infty} m_{Z_{T_n}}.$$

DEPARTMENT OF MATHEMATICS, TULANE UNIVERSTIY, NEW ORLEANS, LA 70118
E-mail address: mdupre@tulane.edu