A Verification of the Central Limit Theorem

Central Limit Theorem: If $\{X_1, X_2, \ldots\}$ is a set of independent and identically distributed random variables with mean μ and standard deviation σ, then the distribution function for the random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the distribution function for the standard normal distribution as $n \to \infty$, where the sample mean random variable is defined as $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

Proof: We start by proving that the *moment generating function* (defined shortly) of the random variable Z_n approaches the moment generating function of the standard normal random variable Z as n approaches infinity. The point is that the moment generating function of a random variable uniquely determines the distribution function of the random variable, although we will not prove this fact. Some details in this prove will not be completely rigorous, since a rigorous proof is beyond the level of this text.

The moment generating function $M_X(t)$ of a random variable X is defined to be the expectation of the random variable e^{Xt}: $M_X(t) = E[e^{Xt}]$, where t can be any real number. Since

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

is the density function for the standard normal random variable Z, then moment generating function of Z is:

(1) $$M_Z(t) = E[e^{Zt}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} e^{zt}\, dz.$$

Completing the square, we have

(2) $$-\frac{1}{2} z^2 + zt = -\frac{1}{2}(z-t)^2 + \frac{1}{2} t^2.$$

Substituting the right side of equation (2) into the right side of equation (1) yields

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-(z-t)^2/2}\, dz.$$

Making the change of variables $u = z - t$ results in

(3) $$M_Z(t) = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-u^2/2} du = e^{t^2/2}.$$

To obtain the final expression on the right side of equation (3), we used the fact that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = 1$$

since $e^{-u^2/2}/\sqrt{2\pi}$ is the density function for the standard normal random variable. We have left to show that $M_{Z_n}(t) \to e^{t^2/2}$ as $n \to \infty$. We can express

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}.$$

Thus, the moment generating function for Z_n is:

(4) $$M_{Z_n}(t) = E\left[e^{\frac{t\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}}\right].$$

Since the $X_i{}'s$ are independent random variables, then Z_n has the density function $p(x_1)p(x_2)\cdots p(x_n)$where $p(x_i)$ is the density function for the random variable X_i. Referring to equation (4), we can express the moment generating function in the form

$$M_{Z_n}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\frac{t\sum_{i=1}^{n}(x_i - \mu)}{\sigma\sqrt{n}}} p(x_1)p(x_2)\cdots p(x_n)dx_1dx_2\cdots dx_n$$

(5) $$= \left[\int_{--\infty}^{\infty} e^{\frac{(x_1-\mu)t}{\sigma\sqrt{n}}} p(x_1)dx_1\right]\left[\int_{--\infty}^{\infty} e^{\frac{(x_2-\mu)t}{\sigma\sqrt{n}}} p(x_2)dx_2\right]\cdots\left[\int_{--\infty}^{\infty} e^{\frac{(x_n-\mu)t}{\sigma\sqrt{n}}} p(x_n)dx_n\right].$$

From Taylor series, we know that $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^2 + \frac{1}{4!}x^2 + \cdots$. Hence, for each $i = 1, 2, \ldots, n$, we have:

(6) $$\int_{-\infty}^{\infty} e^{\frac{(x_i-\mu)t}{\sigma\sqrt{n}}} p(x_i)dx_i = \int_{--\infty}^{\infty}[1 + \frac{(x_i-\mu)t}{\sigma\sqrt{n}} + \frac{(x_i-\mu)^2t^2}{2\sigma^2 n} + \frac{(x_i-\mu)^3t^3}{6\sigma^3 n^{3/2}}\cdots]p(x_i)dx_i$$

Since $p(x_i)$ is the density function for a random variable with mean μ and variance σ^2, then we know

(7) $$\int_{-\infty}^{\infty} p(x_i)\, dx_i = 1, \int_{-\infty}^{\infty} x_i p(x_i)\, dx_i = \mu, and\ \int_{-\infty}^{\infty}(x_i - \mu)^2 p(x_i)\, dx_i = \sigma^2.$$

Although we don't know the precise values of the following integrals, we can say that

(8) $$\int_{-\infty}^{\infty}(x_i - \mu)^k p(x_i)\, dx_i = C(k)$$

for some constants $C(k)$ that are independent of n for k = 3, 4, … Based on equations (7) and (8), we rewrite equation (6) in the form:

(9) $$\int_{-\infty}^{\infty} e^{\frac{(x_i-\mu)t}{\sigma\sqrt{n}}} p(x_i)dx_i = 1 + \frac{t^2}{2n} + \frac{C(3)t^3}{6\sigma^3 n^{3/2}} + \frac{C(4)t^4}{24\sigma^4 n^2} + \cdots.$$

Notice that the right side of equation (9) is the same for every i. Substituting the right side of equation (9) into equation (5) yields

$$M_{Z_n}(t) = (1 + \frac{t^2}{2n} + \frac{C(3)t^3}{6\sigma^3 n^{3/2}} + \frac{C(4)t^4}{24\sigma^4 n^2} + \cdots)^n.$$

Since

$$\frac{C(3)t^3}{6\sigma^3 n^{3/2}} + \frac{C(4)t^4}{24\sigma^4 n^2} + \cdots$$

is much smaller than $\frac{t^2}{2n}$ as $n \to \infty$, then

$$M_{Z_n}(t) \to (1 + \frac{t^2}{2n})^n$$

as $n \to \infty$. We know from calculus that

$$\left(1 + \frac{a}{n}\right)^n \to e^a$$

as $n \to \infty$. Choosing $a=t^2/2$, this shows that

$$M_{Z_n}(t) \to e^{\frac{t^2}{2}}$$

as $n \to \infty$ as we set out to verify.