The Central Limit Theorem

November 19, 2009

Convergence in distribution $X_n \to^{\mathcal{D}} X$ is defined to by

$$\lim_{n\to\infty} Eh(X_n) = Eh(X).$$

or every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$.

However, it is not necessary to verify this for each choice of h. We can limit ourselves to a smaller so-called **convergence determining** family of functions.

- For random variables taking values in the natural numbers, $\{h_z(x) = z^x; |z| < 1\}$ is convergence determining. In this case, we are look at convergence of the probability generating function.
- For real valued random variables, $\{h_t(x) = \exp tx; -h < t < h\}$ is convergence determining provided the necessary expected values exist. Note that $\exp tx$ is not bounded and so we need to make an additional argument to include these function. In this case, we are looking at convergence of the moment generating function.

Example 1. *For the binomial distribution with parameters n and p, the probability generating function is*

$$\rho_{X_n}(z) = ((1-p)+pz)^n = (1-p(1-z))^n$$

If we take the success probability $p = \lambda/n$ to depend on n, then

$$\rho_{X_n}(z) = ((1-p)+pz)^n = \left(1 - \frac{\lambda}{n}(1-z)\right)^n \to \exp\lambda(1-z) = \rho_X(z),$$

the probability generating function for a Poisson random variable X with parameter λ. Thus, we have that the given binomial random variables converges in distribution to a Poisson random variable.

To use this, assume that n is large, but $\lambda = np$ is moderate the binomial random variable is well approximated by a Poisson random. In particular, $Eh(X_n) \approx Eh(X)$ for any bounded continuous h

1 Central Limit Theorem

If we look at distributions for the sum $T_n = X_1 + X_2 + \cdots + X_n$, what do we see. Let's look first to the simplest case, X_i Bernoulli random variables.

This is looking like the bell curve. To make the comparisons fair, let look at standardized versions of the random variables with mean μ and variance σ^2,

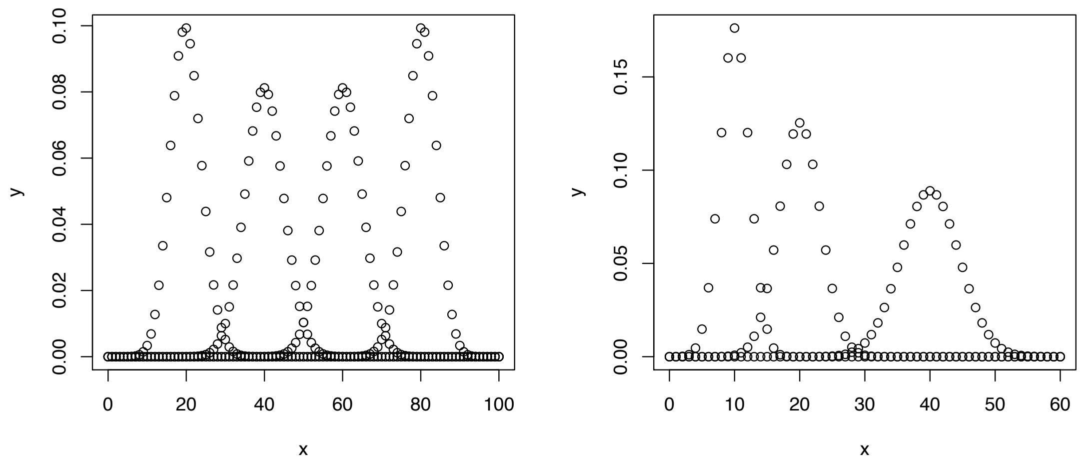

Figure 1: a. Successes in 100 Bernoulli trials with $p = 0.2, 0.4, 0.6$ and 0.8. b. Successes in Bernoulli trials with $p = 1/2$ and $n = 20, 40$ and 80.

$$Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} \tag{1}$$

and look at the density of the sum of standardized exponential random variables.

Again, we see the densities approaching that of a bell curve. The **classical central limit theorem** states that if $\{X_i; i \geq 1\}$ and independent and identically distributed with common mean μ and common variance σ^2, then Z_n as defined by equation (**??**) converges to Z a standard normal random variable. In terms of the cumulative distribution function

$$\lim_{n\to\infty} P\{Z_n \leq z\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx = \Phi(z)$$

where Φ is the cumulative distribution function of the standard normal.

We will prove this in the case that the X_i have a moment generating function $M_X(t)$ for the interval $t \in (-h, h)$ by showing that

$$\lim_{n\to\infty} M_{Z_n}(t) = \exp\frac{t^2}{2}$$

or equivalently, show that the cumulant generating functions $K_{Z_n}(t) = \log M_{Z_n}(t)$ satisfy

$$\lim_{n\to\infty} K_{Z_n}(t) = \frac{t^2}{2}$$

Write

$$Y_i = \frac{X_i - \mu}{\sigma}$$

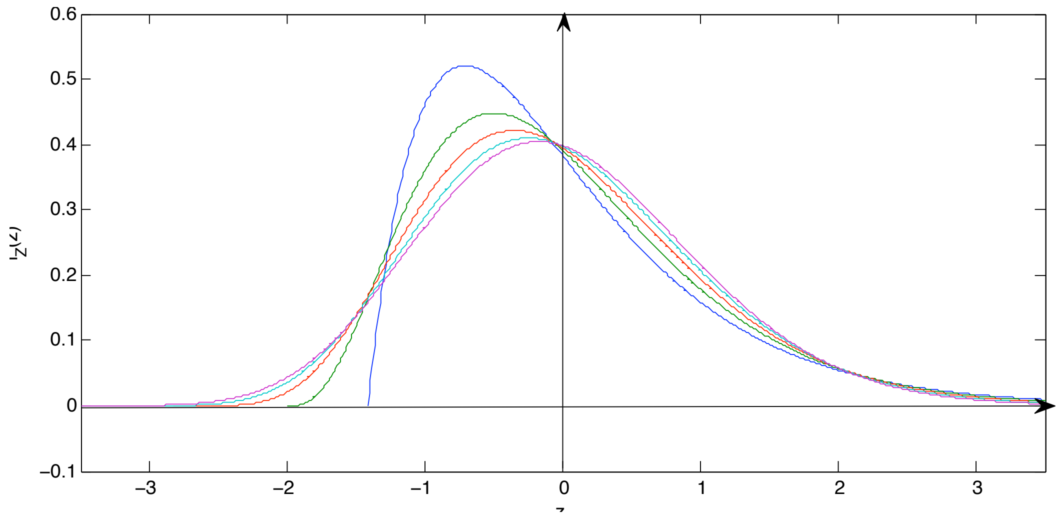

Figure 2: Density of the standardized version of the sum of n independent exponential random variables for $n = 2, 4, 8, 16$ and 32.

then

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

For $M_Y(t)$ the moment generating function for the Y_i and $M_{T_n}(t)$, the moment generating function for T_n,

$$M_{Z_n(t)} = E\exp(tZ_n) = E\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right) = \prod_{i=1}^{n} E\exp\left(\frac{t}{\sqrt{n}} Y_i\right) = \left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

and

$$K_{Z_n}(t) = n \log M_Y\left(\frac{t}{\sqrt{n}}\right) = nK_Y\left(\frac{t}{\sqrt{n}}\right).$$

Recall that for the cumulant generating function K_Y,

$$K_Y'(0) = EY_1 = 0, \quad K_Y''(0) = \mathrm{Var}(Y) = 1.$$

Finally, from two applications of L'Hôpital's rule,

$$\lim_{n\to\infty} K_{Z_n}(t) = \lim_{n\to\infty} nK_Y\left(\frac{t}{\sqrt{n}}\right) = \lim_{\epsilon\to 0} \frac{K_Y(\epsilon t)}{\epsilon^2} = \lim_{\epsilon\to 0} \frac{tK_Y'(\epsilon t)}{2\epsilon} = \lim_{\epsilon\to 0} \frac{t^2 K_Y''(\epsilon t)}{2}. = \frac{t^2 K_Y''(0)}{2} = \frac{t^2}{2}.$$

Example 2. *For Bernoulli trials, $\mu = p$ and $\sigma^2 = p(1-p)$. Thus, for large enough n*

$$Z_n = \frac{T_n - np}{\sqrt{np(1-p)}},$$

has approximately the distribution of a standard normal random variable. For 100 tosses of a fair coin,

$$Z_n = \frac{T_n - 50}{5},$$

and

$$\{T_n \le 40\} = \{Z_n \le -2\}$$

So,

$$P\{T_n \le 40\} \approx P\{Z \le -2\} = 0.054.$$

Example 3. *For an exponential sample with mean 1. Then, the standard deviation is also 1 and for 64 observations*

$$Z_n = \frac{T_n - 64}{8},$$

$$\{T_n \ge 78\} = \{Z_n \ge 1.75\}$$

So,

$$P\{T_n \ge 78\} \approx P\{Z \ge -1.75\} = 0.086.$$

2 Slutsky's Theorem

Some useful extensions of the central limit theorem are based on Slutsky's theorem.

Theorem 4. *Let $X_n \to^{\mathcal{D}} X$ and $Y_n \to^{P} a$, a constant as $n \to \infty$. Then*

1. $Y_n X_n \to^{\mathcal{D}} aX$, *and*
2. $X_n + Y_n \to^{\mathcal{D}} X + a$.

For example, by the law of large numbers, the sample variance

$$S_n^2 \to^{a.s.} \sigma^2,$$

the distribution variance as $n \to \infty$. Thus,

$$\frac{S_n}{\sigma} \to^{a.s.} 1.$$

Thus, it also converges in probability. So, by Slutsky's theorem, the t-statistic

$$\frac{T_n - n\mu}{S_n\sqrt{n}} = \frac{S_n}{\sigma}\frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n}{\sigma} Z_n \to^{\mathcal{D}} 1 \cdot Z,$$

a standard normal as $n \to \infty$

3 Delta Method

For a random sample $\{X_n \geq 1\}$ with common mean μ and common variance σ^2, we can write the central limit theorem using the sample mean.

$$\sqrt{n}(\bar{X}_n - \mu) \ \to^{\mathcal{D}} \sigma Z$$

where Z is a standard normal.

To generalize this, assume that $\{Y_n \geq 1\}$ is a sequence of random variables satisfying

$$\sqrt{n}(Y_n - \theta) \to^{\mathcal{D}} \sigma Z$$

for some value θ

Then the **delta method** states that if a function g has a continuous derivative and $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(Y_n) - g(\theta)) \ \to^{\mathcal{D}} \sigma g'(\theta)\tilde{Z}$$

where $\tilde{Z}$ is also a standard normal.

To prove this, expand g is a Taylor's series about the value θ

$$g(Y_n) = g(\theta) + g'(\tilde{\theta})(Y_n - \theta),$$

or

$$\sqrt{n}(g(Y_n) - g(\theta)) = g'(\tilde{\theta})\sqrt{n}(Y_n - \theta).$$

where $\tilde{\theta}$ lies between Y_n and θ. Note that since $Y_n \to^P \theta$ implies $\tilde{\theta} \to^P, \theta$ and $g'(\theta)$ is continuous,

$$g'(\tilde{\theta}) \to^P g'(\theta).$$

and the theorem follows from applying Slutsky's theorem.

Example 5. *For Bernoulli trials, write $\bar{X} = \hat{p}$, then*

$$\sqrt{n}(\hat{p} - p) \to^{\mathcal{D}} \sqrt{p(1-p)}Z.$$

If we could find g so that

$$g'(p) = \frac{1}{\sqrt{p(1-p)}},$$

then

$$\sqrt{n}(g(\hat{p}) - g(p)) \to^{\mathcal{D}} Z.$$

Such a choice, which here is $g(p) = 2\arcsin(\sqrt{p})$ is called a **variance stabilizing transformation***.*