[7.3] The Central Limit Theorem!

$\boxed{\text{Thm } 7.3.1}$ Let $Y_1, Y_2, \cdots$ be a sequence of RV's with CDFs $G_1(y), G_2(y), \ldots$ and with MGFs $M_1(y), M_2(y), \ldots$

If $M(t)$ is the MGF of a CDF $G(y)$, and if

$$\lim_{n\to\infty} M_n(t) = M(t) \text{ for all } t \in (-h,h)$$

Then $\lim_{n\to\infty} G_n(y) = G(y)$ for all points of continuity of $G(y)$

In other words, $Y_n \xrightarrow{d} Y$ if $M_n(t) \to M(t)$

Preliminary

We will make use of these limits several times:

$$\boxed{\begin{aligned} &\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^{nb} = e^{cb} \\ &\lim_{n\to\infty}\left(1+\frac{c}{n}+\frac{d(n)}{n}\right)^{nb} = e^{cb} \quad \left(\text{If } d(n)\to 0 \text{ as } n\to\infty\right)\end{aligned}}$$

Ex: Let $X_1, \dots, X_n$ be a RS from $X_i \sim BIN(1, p)$

Consider $Y_n = \sum_{i=1}^{n} X_i$. Suppose $n \to \infty$ and $p \to 0$ such that $\mu = np$ is fixed $\mu > 0$

It follows that

$$M_n(t) = (pe^t + q)^n = \left[\frac{\mu e^t}{n} + 1 - \frac{\mu}{n}\right]^n$$

(Note: $p = \frac{\mu}{n}$, $q = 1 - p$)

$$= \left[1 + \frac{\mu(e^t - 1)}{n}\right]^n \longrightarrow \text{so } c = \mu(e^t - 1)$$

Thus, $\lim\limits_{\substack{n \to \infty \\ p \to 0 \\ \mu = np}} M_n(t) = \lim \left[1 + \frac{c}{n}\right]^n = e^c = e^{\mu(e^t - 1)} = M(t)$

$M(t)$ is identified as MGF for the Poisson. Thus $Y_n \xrightarrow{d} Y \sim POI(\mu)$

Continuing this example, suppose instead of Y_n, we choose $W_n = \frac{Y_n}{n}$, the sample proportion. $Y_n = \sum_{i=1}^{n} X_i$

The series expansion for e^x is $\sum_{k=0}^{\infty} \frac{x^k}{k!}$, so

$\boxed{\text{Bernoulli Law of Large Numbers}}$

$$M_{W_n}(t) = (pe^{t/n}+q)^n = \left(p\sum_{k=0}^{\infty}\frac{(t/n)^k}{k!}\right) + q)$$

$$= \left[p\left(1+\frac{t}{n}+\frac{t^2}{2n^2}+\frac{t^3}{3!n^3}+\cdots\right)+q\right]$$

$$= \Big[\underbrace{p+q}_{1} + \frac{pt}{n} + \underbrace{\frac{pt^2}{2n^2} + \frac{pt^3}{3!n^3} + \cdots}_{\text{factor } \frac{1}{n} \text{ out.}}\Big]^n$$

$$= \left[1+\frac{pt}{n}+\underbrace{\frac{1}{n}\left(\frac{pt^2}{2n}+\frac{pt^3}{3!n^2}+\cdots\right)}_{d(n)\to 0}\right]^n = \left[1+\frac{pt}{n}+\frac{d(n)}{n}\right]^n \to e^{pt}$$

Can you identify the MGF?

Yep, it's the degenerate dist with p for its parameter.

Therefore $W_n = \frac{Y_n}{n} \xrightarrow{d} DEG(p)$ or $W_n \xrightarrow{P} p$

W_n converges stochastically to p, the prob of success

Ex: Now consider the sequence of $Z_n = \frac{Y_n - np}{\sqrt{npq}}$

Let $\sigma^2 = npq$. Then $Z_n = \frac{Y_n}{\sigma} - \frac{np}{\sigma}$

$$M_{Z_n}(t) = E\left[e^{tZ_n}\right] = E\left[e^{-\frac{npt}{\sigma} + \frac{tY_n}{\sigma}}\right] = e^{-\frac{npt}{\sigma}} \underbrace{E\left[e^{\left(\frac{t}{\sigma}\right)Y_n}\right]}_{M_{Y_n}\left(\frac{t}{\sigma}\right)} = e^{-\frac{npt}{\sigma}}\left(pe^{\frac{t}{\sigma}} + q\right)^n$$

$$= \Big[\underbrace{e^{-\frac{pt}{\sigma}}}_{①}\underbrace{\left(pe^{t/\sigma} + q\right)}_{②}\Big]^n$$

⇒ Do separately... Expand both in a power series and then multiply together on the next page!

Since $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = 1 + x + \frac{x^2}{2} + o(x^3)$

Then

① $$e^{-\frac{pt}{\sigma}} = 1 - \frac{pt}{\sigma} + \frac{p^2t^2}{2\sigma^2} + o(t^3)$$

② $$pe^{\frac{t}{\sigma}} + \underbrace{q}_{1-p} = p\left[1 + \frac{t}{\sigma} + \frac{t^2}{2\sigma^2} + o(t^3)\right] + \underbrace{q}_{1-p}$$

$$= \cancel{p} + \frac{pt}{\sigma} + \frac{pt^2}{2\sigma^2} + 1 - \cancel{p} + o(t^3)$$

$$= 1 + \frac{pt}{\sigma} + \frac{pt^2}{2\sigma^2} + o(t^3)$$

So $$e^{-\frac{pt}{\sigma}}\left(pe^{\frac{t}{\sigma}} + q\right) = \left(\underset{a_0}{1} - \underset{a_1}{\frac{pt}{\sigma}} + \underset{a_2}{\frac{p^2t^2}{2\sigma^2}} + o(t^3)\right)\left(\underset{b_0}{1} + \underset{b_1}{\frac{pt}{\sigma}} + \underset{b_2}{\frac{pt^2}{2\sigma^2}} + o(t^3)\right)$$

Note: We need to multiply it out (only part) (Keep terms better than $O(t^3)$)

Multiplying

$$\Big[\underbrace{1}_{a_0} - \underbrace{\frac{pt}{\sigma}}_{a_1} + \underbrace{\frac{p^2t^2}{2\sigma^2}}_{a_2} + o(t^3)\Big]\Big[\underbrace{1}_{b_0} + \underbrace{\frac{pt}{\sigma}}_{b_1} + \underbrace{\frac{pt^2}{2\sigma^2}}_{b_2} + o(t^3)\Big]$$

$$= a_0 b_0 + (a_1 b_0 + a_0 b_1) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + o(t^3)$$

$$= 1 + \cancel{\left(-\frac{pt}{\sigma} + \frac{pt}{\sigma}\right)} + \left(\frac{pt^2}{2\sigma^2} - \frac{2p^2t^2}{2\sigma^2} + \frac{p^2t^2}{2\sigma^2}\right) + o(t^3)$$

$p(1-p) = pq$

$$= 1 + \left(\frac{p-p^2}{2\sigma^2}\right)t^2 + o(t^3) = 1 + \frac{pq}{2\sigma^2}t^2 + o(t^3)$$

Since $\sigma^2 = npq$, then $\frac{pq}{2\sigma^2} = \frac{\cancel{pq}}{2n\cancel{pq}} = \frac{1}{2n}$, then

$$= 1 + \frac{(t^2/2)}{n} + o\left(\frac{1}{n^3}\right) = 1 + \frac{(t^2/2)}{n} + \frac{d(n)}{n}$$

$o(\frac{1}{n^2}) \to 0$

Thus, the limit is $\lim\limits_{n\to\infty} M_{Z_n}(t) = \lim\limits_{n\to\infty}\left[1 + \frac{t^2/2}{n} + \frac{d(n)}{n}\right]^n = e^{t^2/2}$

this is the MGF for the standard Normal! $Z_n \xrightarrow{d} Z \sim N(0,1)$

Thm 7.3.2 Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a RS from a dist with mean μ and variance $\sigma^2 < \infty$, then the limiting dist of

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

is the standard normal dist. Or simply

$$Z_n \xrightarrow{d} Z \sim N(0,1) \text{ as } n \to \infty$$

Proof: The result holds for any dist with finite mean and variance, but the proof is easier by assuming the MGF exists!

(If the MGF doesn't exist, the characteristic function does. It is a function of complex variables, so lets use the MGF proof: Let $m(t)$ denote the MGF of $X-\mu$.

Then: $m(t) = M_{X-\mu}(t)$ and $m(0)=1$, $m'(0) = E(X-\mu) = E(X)-\mu = 0$

$m''(0) = E(X-\mu)^2 = \sigma^2$

So, expanding $m(t)$ in a Taylor series about 0 gives

$$m(t) = m(0) + m'(0)t + m''(\xi)\frac{t^2}{2}, \text{ where } \xi \in (0,t)$$

Z_n can also be written as $\dfrac{\sum X_i - n\mu}{\sigma\sqrt{n}} = \dfrac{\sum_1^n X_i - \sum_1^n \mu}{\sigma\sqrt{n}} = \dfrac{\sum_1^n (X_i-\mu)}{\sigma\sqrt{n}}$

So the MGF of Z_n is:

$$M_{Z_n}(t) = E\left[\exp\left\{\frac{\sum_1^n (X_i-\mu)}{\sigma\sqrt{n}}t\right\}\right] = E\left[\prod_{i=1}^n \frac{(X_i-\mu)t}{\sigma\sqrt{n}}\right] = \prod_{i=1}^n E\left[(X_i-\mu)\frac{t}{\sigma\sqrt{n}}\right]$$

$$= \prod_{i=1}^n M_{X_i-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right) \quad \text{(Note: } M_{X_i-\mu}(t) = m(t) \text{)} \quad X_i \sim iid$$

$$= \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Continuing,

$$M_{z_n}(t) = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 + \frac{m''(\xi)\,t^2}{n\sigma^2}\right]^n, \quad 0 < |\xi| < \frac{|t|}{\sigma\sqrt{n}}$$

Now, since $\xi \in \left(\frac{-t}{\sigma\sqrt{n}}, \frac{t}{\sigma\sqrt{n}}\right)$, then $\xi \longrightarrow 0$ as $n \to \infty$

Hence, $m''(\xi) = m''(0) = \sigma^2$

$$M_{z_n}(t) = \left[1 + \frac{\sigma^2 t^2}{2n\sigma^2} - \frac{\sigma^2 t^2}{2n\sigma^2} + \frac{m''(\xi)\,t^2}{2n\sigma^2}\right]^n$$

Add zero trick

$$= \left[1 + \frac{(t^2/2)}{n} + \frac{(m''(\xi) - \sigma^2)t^2}{2n\sigma^2}\right]^n$$

Let $d(n) = \dfrac{m''(\xi) - \sigma^2)}{2\sigma^2}$, Note $d(n) \to 0$ as $n \to \infty$, so the limit is

Thus, $\lim\limits_{n\to\infty} M_{z_n}(t) = e^{t^2/2} = M_z(t)$ Thus, $z_n \xrightarrow{d} z \sim N(0,1)$

Ex: Roll 20 dice. (6 sided) $X_i \sim DU(6)$

Find $P\left[\sum_1^n X_i > 85\right] \Rightarrow \mu = E(X_i) = \frac{N+1}{2} = \frac{7}{2}$

$$\sigma^2 = V(X_i) = \frac{N^2-1}{12} = \frac{35}{12}$$

Remember, $Z_n = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$

So $P\left[\sum X_i > 85\right] \doteq P\left[Z > \frac{85 - 20\left(\frac{7}{2}\right)}{\sqrt{20}\sqrt{\frac{35}{12}}}\right] = P\left[Z > 1.96\right] = 0.0250$

So the prob is 2.5% to get a sum of 85!