Stat 401 - 12/1/17

Sec 5.3 - CLT (Central Limit Thm)

Thm) Let $X_1, X_2, \ldots, X_n$ denote the observations of a random sample from a distribution that has mean μ and variance $\sigma^2 < \infty$.

Then
$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{D} N(0,1)$$

Why do we care?

- X_i's don't have to start out Normal.
- helps to calculate different prob.
- help determine how large a sample size n needs to be to make sure that our calculations will give us what we want.

Proof of CLT

Suppose $M_X(t) = \mathbb{E}[e^{tX}]$

Define $m(t) = M_{X-\mu}(t)$

$$= \mathbb{E}[e^{t(X-\mu)}]$$

$$= \mathbb{E}[e^{tX} e^{-\mu t}]$$

$$= e^{-\mu t} \mathbb{E}[e^{tX}]$$

$$= e^{-\mu t} M_X(t)$$

Them $m(0) = e^{-\mu(0)} M_X(0)$

$$= (1)\,\mathbb{E}[e^{0X}]$$

$$= (1)\,\mathbb{E}[1]$$

$$= 1$$

$$m'(0) = \mathbb{E}[X-\mu] = \mathbb{E}[X] - \mu$$

$M_X'(t) = \mathbb{E}[X]$

$$= \mu - \mu$$

$$= 0$$

$$m''(0) = E[(X-\mu)^2] = \cancel{E[X^2] - (E[X])^2}$$
$$= Var(X)$$
$$= \sigma^2$$

By Taylor's Formula (Expansion) there exists a ~~ee~~ value c that is between 0 and t such that

$$m(t) = m(0) + \underbrace{m'(0)}_{=0}t + \frac{m''(c)t^2}{2}$$
$$= 1 + \frac{m''(c)t^2}{2}$$

$$= 1 + \frac{\sigma^2 t^2}{2} - \frac{\sigma^2 t^2}{2} + \frac{m''(c)t^2}{2}$$

$$= 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(c) - \sigma^2]t^2}{2}$$

Look at $Y_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$

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

Find $M_{Y_n}(t)$

$$M_{Y_n}(t) = \mathbb{E}\left[e^{t\left(\frac{\sum X_i - n\mu}{\sigma\sqrt{n}}\right)} \right]$$

indep.

$$= \mathbb{E}\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)} \right] \cdot \ldots \cdot \mathbb{E}\left[e^{t\left(\frac{X_n-\mu}{\sigma\sqrt{n}}\right)} \right]$$

$$n\mu = \mu + \underbrace{\ldots + \mu}_{n \text{ times}}$$

$$= \left(\mathbb{E}\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)} \right] \right)^n$$

$$= \left(\mathbb{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)} \right] \right)^n$$

$$= \left(M_{X_1-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n = \left(m\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n$$

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

$$= 1 + \frac{t^2}{2n} + \frac{[m''(c)-\sigma^2]\,t^2}{2n\sigma^2}$$

$$\text{for } 0 < c < \frac{t}{\sigma\sqrt{n}}$$

Then

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

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

Recall:

$$\lim_{n\to\infty}\left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn}$$

$$= \lim_{n\to\infty}\left[1 + \frac{b}{n}\right]^{cn} = e^{bc}$$

where $\lim_{n\to\infty} \psi(n) = 0$

Note: $m''(t)$ is continuous at $t=0$.

so $\lim_{n\to\infty} [m''(c) - \sigma^2] = 0$.

as $c \to 0$

Thus

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

$$\xrightarrow{\text{as } n\to\infty} e^{\frac{t^2}{2}(1)}$$

$$= e^{0t + \frac{1}{2}(1)^2 t^2}$$

$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ = MGF for $N(\mu, \sigma^2)$

$\rightarrow$ MGF for $N(0, 1)$

$\square$

Sample Size.

ex) Application.

Suppose we have Bernoulli Trials with prob p.

Recall : $\mu = p \qquad \sigma^2 = p(1-p)$

We know $\bar{X}_n \xrightarrow{P} \mu = p$

How large does n have to be in order for

$$\mathbb{P}\left\{ \left| \bar{X}_n - p \right| < 0.01 \right\} \geq 99\%)$$

$\boxed{\text{SOL}}$ By the CLT:

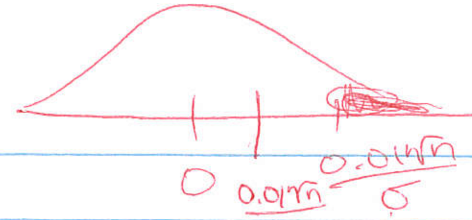
$$\frac{\sqrt{n}\,(\bar{X}_n - p)}{\sigma} \sim N(0,1)$$

$$\mathbb{P}\left\{ \left| \bar{X}_n - p \right| < 0.01 \right\}$$

$$= \mathbb{P}\left\{ \left| \frac{\sqrt{n}\,(\bar{X}_n - p)}{\sigma} \right| < \frac{0.01\sqrt{n}}{\sigma} \right\}$$

$$\simeq \mathbb{P}\left\{ \left| N(0,1) \right| < \frac{0.01\sqrt{n}}{\sigma} \right\}$$

want $\geq$ 99% or 0.99



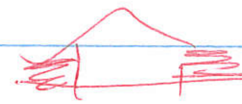
i.e. want

$$\left[P\left\{ |N(0,1)| > \frac{0.01\sqrt{n}}{\sigma} \right\} < 0.01 \right]$$

since $\sigma^2 = p(1-p) < 1$

since $0 < p < 1$

$\Rightarrow \sigma < 1$



$$< \left[P\{ |N(0,1)| > 0.01\sqrt{n} \} < 0.01 \right]$$

$$= \left[2P\{ N(0,1) > 0.01\sqrt{n} \} < 0.01 \right]$$

$$= \left[P\{ N(0,1) > 0.01\sqrt{n} \} < 0.005 \right]$$

$\Rightarrow$ want

$$P\{ Z \le 0.01\sqrt{n} \} \ge 0.995$$

$$z = 2.58 \le 0.01\sqrt{n}$$

$$258 < \sqrt{n}$$

$$66564 = 258^2 < n$$