1) Large Sample theory is used to show statistical methods work.

2) Know p73 Central Limit Theorem (CLT)

Let $Y_1, \ldots, Y_n$ be iid with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 > 0$. Then

$$\sqrt{n}\,(\overline{Y}_n - \mu) \underbrace{\xrightarrow{D}}_{\approx \text{ for now}} N(0, \sigma^2).$$

$$\text{so } \underbrace{\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right)}_{z \text{ score of } \overline{Y}_n} = \underbrace{\sqrt{n}\left(\frac{\sum_{i=1}^n Y_i - n\mu}{n\sigma}\right)}_{z \text{ score of } \sum_{i=1}^n Y_n} \xrightarrow{D} N(0,1)$$

3) Know 2 applications i) $\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$,

ii) Suppose $Y_n = \sum_{i=1}^n X_i$ where the X_i are iid with mean μ_X and variance σ_X^2.

Then $\sqrt{n}(\overline{X}_n - \mu_X) = \sqrt{n}\left(\frac{Y_n}{n} - \mu_X\right) \xrightarrow{D} N(0, \sigma_X^2)$.

4) Th a) If $Y_1, \ldots, Y_n$ are ind $BIN(k_i, \rho)$ RVS, then $\sum_{i=1}^n Y_i \sim BIN\left(\sum_{i=1}^n k_i, \rho\right)$.

b) Denote a χ^2_p RV by $\chi^2(p)$. If $Y_1, \ldots, Y_n$ are ind $\chi^2_{p_i}$, then

$$\sum_{i=1}^n Y_i \sim \chi^2\left(\sum_{i=1}^n p_i\right).$$

c) If $Y_1, \ldots, Y_n$ are iid $EXP(\lambda)$, then $\sum_{i=1}^n Y_i \sim G(n, \lambda)$.

d) If $Y_1, \ldots, Y_n$ are ind Gamma $G(\nu_i, \lambda)$, then $\sum_{i=1}^n Y_i \sim G\left(\sum_{i=1}^n \nu_i, \lambda\right)$.

Hence if $Y_1, \ldots, Y_n$ are iid $G(\nu, \lambda)$, then

$$\sum_{i=1}^n Y_i \sim G(n\nu, \lambda).$$

e) If $Y_1, \ldots, Y_n$ are ind $N(\mu_i, \sigma_i^2)$ then $\sum_{i=1}^n (a_i + b_i Y_i) \sim N\left(\sum_{i=1}^n (a_i + b_i \mu_i), \sum_{i=1}^n b_i^2 \sigma_i^2\right)$ where a_i and b_i are constants.

If $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$,

then $\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

f) If $Y_1, \ldots, Y_n$ are ind $Pois(\theta_i)$,

then $\sum_{i=1}^{n} Y_i \sim Pois\left(\sum_{i=1}^{n} \theta_i\right)$.

Thus if $Y_1, \ldots, Y_n$ are iid $Pois(\theta)$,

then $\sum_{i=1}^{n} Y_i \sim Pois(n\theta)$.

This theorem can be proved with mgfs or characteristic functions.

5] know p86 Delta Method

If $g'(\theta) \neq 0$ and $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$

where $\sigma^2 = \sigma^2(\theta)$, then

$$\sqrt{n}\left(g(T_n) - g(\theta)\right) \xrightarrow{D} N\left(0, \sigma^2 [g'(\theta)]^2\right).$$

Know p89-90 2.5

6) <u>Second Order Delta method</u>

Suppose $g'(\theta) = 0$, $g''(\theta) \neq 0$

and $\sqrt{n}\,(T_n - \theta) \xrightarrow{D} N(0, \tau^2(\theta))$.

Then $n[\,g(T_n) - g(\theta)\,] \xrightarrow{D}$

$\frac{1}{2}\tau^2(\theta)\, g''(\theta)\, \chi^2_1$. Could use $\sigma^2(\theta)$.

7) RHS in 2), 3), 5) and 6)

does <u>not</u> depend on n.

8) problem will often have

a) CLT

b) Delta method

c) second order delta method, $\Big\}$ less often

9) For CLT, a) sometimes you

need to find $E(X)$ and $E(X^2)$.

Then $V(X) = E[X^2] - [E(X)]^2$.

variant: given formula for $E[Y^r]$, find $E(Y) = E(Y^1)$ with $r=1$ and $E(Y^2)$ with $r=2$.

drive online text soln in back

ex) see old Q1, 02.39a, 02.12, 02.22a) 02.26, 02.27, 02.28b, 02.33a) d) 0.2.34a), 0.2.35ab, 02.36ab 02.37, 02.38

ex) Suppose $X_1, \ldots, X_n$ are iid with $E(X^k) = \frac{2\theta^k}{k+2}$. Find the limiting dist of $\sqrt{n}(\bar{X}_n - c)$ for appropriate c.

soln) $E(X) = \frac{2\theta}{3}$, $E(X^2) = \frac{\theta^2}{2}$

$V(X) = E(X^2) - [E(X)]^2 = \frac{\theta^2}{2} - \frac{4\theta^2}{9}$

$= \frac{9\theta^2 - 8\theta^2}{18} = \frac{\theta^2}{18}$. So

$\sqrt{n}\left(\bar{X}_n - \frac{2\theta}{3}\right) \xrightarrow{D} N\left(0, \frac{\theta^2}{18}\right)$ by CLT.

ex) Let $X_1, \ldots, X_n$ be iid Bernoulli(p) RVs.

Let $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

a) Find the asymptotic distribution of $\sqrt{n}(Y_n - p)$ when $p = 0.3$.

b) Let $g(Y_n) = Y_n(1-Y_n)$. Find the asymptotic distribution of $\sqrt{n}(g(Y_n) - g(p))$ when $p = 0.3$.

c) Let $g(Y_n) = Y_n(1-Y_n)$. Find the asymptotic distribution of $n(g(Y_n) - g(p))$ when $p = 0.5$.

soln.. a) $E(X_i) = p$, $V(X_i) = p(1-p)$

$$\sqrt{n}(Y_n - p) = \sqrt{n}(\overline{X}_n - p) \xrightarrow{D} N(0, p(1-p))$$

by CLT. $.3(.7) = 0.21$, so

$$\boxed{\sqrt{n}(Y_n - 0.3) \xrightarrow{D} N(0, 0.21)}$$

b) Let $g(x) = x(1-x) = x - x^2$.

So $g'(x) = 1 - 2x$ and $[g'(p)]^2 = (1-2p)^2$.

So $\sqrt{n}\,[Y_n(1-Y_n) - p(1-p)] \xrightarrow{D}$

$N[0, \underbrace{p(1-p)}_{0.21}\,\underbrace{(1-2p)^2}_{0.16}]$ and

$$\boxed{\sqrt{n}\,[Y_n(1-Y_n) - 0.21] \xrightarrow{D} N(0, 0.0336)}$$

by the Delta Method.

c) If $g'(p) = 0$ and $g''(p) \neq 0$,

$n[g(Y_n) - g(p)] \xrightarrow{D} \frac{1}{2}\sigma_p^2\, g''(p)\, \chi_1^2$

by the Second Order Delta method

$g'(0.5) = 0$, $g''(x) = g'(0.5) = -2$

$\sigma_p^2 = \sigma_{\frac{1}{2}}^2 = p(1-p)\big|_{p=0.5} = \frac{1}{2}\,\frac{1}{2} = \frac{1}{4}$

So $\boxed{n\,[Y_n(1-Y_n) - \frac{1}{4}] \xrightarrow{D} -\frac{1}{4}\chi_1^2}$

ex) θ is a dummy variable,

Let $Y_1, \ldots, Y_n$ be iid with $0<\theta<1$,

$E(Y) = \dfrac{r(1-\theta)}{\theta}$, $V(Y) = r\dfrac{(1-\theta)}{\theta^2}$.

a) Find the limiting dist of

$$\sqrt{n}\left(\bar{Y} - \frac{r(1-\theta)}{\theta}\right).$$

b) Let $g(\bar{Y}) = \dfrac{r}{r+\bar{Y}}$. Find the

limiting dist of $\sqrt{n}\,(g(\bar{Y}) - c)$

for appropriate constant c.

Soln a) $\sqrt{n}\left(\bar{Y} - \dfrac{r(1-\theta)}{\theta}\right) \xrightarrow{D} N\left(0, \dfrac{r(1-\theta)}{\theta^2}\right)$

by CLT.

b) Let $\mu = \dfrac{r(1-\theta)}{\theta}$. So $\sqrt{n}(\bar{Y}-\mu) \xrightarrow{D}$

$N(0, \sigma^2)$ with $\sigma^2 = \frac{r(1-\theta)}{\theta^2}$.

By the Delta method,

$$\sqrt{n}\,(g(\bar{Y}) - g(\mu)) \xrightarrow{D} N(0, [g'(\mu)]^2 \sigma^2)$$

$$g(\mu) = \frac{r}{r+\mu} = \frac{r}{r + \frac{r(1-\theta)}{\theta}} =$$

$$\frac{r\theta}{r\theta + r(1-\theta)} = \theta = c.$$

$$g'(\mu) = \frac{d}{d\mu}\, r(r+\mu)^{-1} = \frac{-r}{(r+\mu)^2}$$

$$= -\frac{r}{\left[r + \frac{r(1-\theta)}{\theta}\right]^2} = \frac{-r\theta^2}{(r\theta + r - r\theta)^2}$$

$$= -\frac{\theta^2}{r}. \quad \text{So } [g'(\mu)]^2 = \frac{\theta^4}{r^2}.$$

so $$\boxed{\sqrt{n}\left(g(\bar{Y}) - \theta\right) \xrightarrow{D} N\left(0, \frac{r(1-\theta)}{\theta^2}\frac{\theta^4}{r^2}\right) = N\left[0, \frac{\theta^2(1-\theta)}{r}\right].}$$

ex} Sometimes you need to find g.
see old Q 1b), HW1 c).

10}* The g in the Delta method can not depend on n, since g_n is a sequence of functions rather than a single function.

11} Know p66 Def: Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of RVs with cdfs F_n, and let X be a RV with cdf F. Then Z_n converges in distribution to X

written $X_n \xrightarrow{D} X$, if

$\lim_{n\to\infty} F_n(t) = F(t)$ at each continuity point t of F. The dist X is called the <u>limiting distribution</u> or <u>asymptotic distribution</u> of Z_n.

12) <u>Know</u> The limiting distribution X does not depend on n.

A common error is to get a "limiting distribution" that does depend on n.

13) Know $F(x) = P(X \le x)$.
$0 \le F_n(t) \le 1$ so
$\lim_{n\to\infty} F_n(t) = H(t)$ has $0 \le H(t) \le 1$ if the limit exists.

Common error! get
$H(t) < 0$ or $H(t) > 1$.

14) Know! convergence in distribution is also known as weak convergence and convergence in law, denoted by $Z_n \xrightarrow{L} X$.

15) Know If $F_n(t) \to H(t)$ and $H(t)$ is continuous, then $H(t)$ needs to be a cdf: $H(t) = F_X(t)$ for some RV X if $Z_n \xrightarrow{D} X$.
If $H(t)$ is a constant, then $H(t)$ is not a cdf, and Z_n does not converge in dist to any RV X.