7.2 Sampling Dist. Related to the Normal Dist.

$\boxed{\text{Simulations}}$

Thm 7.1

Let $Y_1 \cdots Y_n$ be a r.s. of size n from a normal dist. with mean $= \mu$ and variance $= \sigma^2$. Then

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \sim N\left(\mu_{\bar{Y}} = \mu,\ \sigma^2_{\bar{Y}} = \sigma^2/n\right)$$

Proof:

Using Thm 6.3 .. $\bar{Y} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n$, $Y_i \sim N(\mu, \sigma^2)$

$\Rightarrow \bar{Y}$, a linear combination of ind. normal RVs has a normal dist.

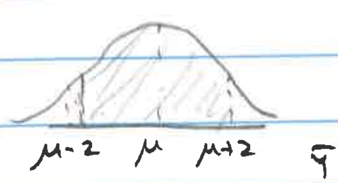
$\Rightarrow E(\bar{Y}) = \frac{1}{n}E(Y_1) + \cdots + \frac{1}{n}E(Y_n) = \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = \frac{n\mu}{n} = \mu$

$\Rightarrow var(\bar{Y}) = \frac{1}{n^2}var(Y_1) + \cdots + \frac{1}{n^2}var(Y_n) = \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$ ∎

7.11 • Forester wants to estimate the avg. basal area of pine trees.

- From past experience he knows these are normally dist. w/ std. dev. $\sigma \approx 4$ inches.
- If he samples 9 trees find $P(|\bar{Y} - \mu| < 2)$

$\bar{Y} \sim N\left(\mu, \sigma_{\bar{Y}} = 4/\sqrt{9}\right)$



$z = \frac{2}{4/3} = 1.5$

$P(|\bar{Y} - \mu| < 2) = P(-1.5 < Z < 1.5) = 2(.4332) = .8664$

7.12 Spse the forester wants $\bar{Y}$ to be within 1 sq. in. of the pop. mean w/ prob. = .90. How many trees need to be selected?

$$P(|\bar{Y}-\mu|<1) = .90 \qquad \sigma_{\bar{Y}} = \frac{4}{\sqrt{n}}$$

$$P\left(\frac{-1}{4/\sqrt{n}} < Z < \frac{1}{4/\sqrt{n}}\right) = .90 \quad \Rightarrow \quad \frac{\sqrt{n}}{4} = 1.645$$

$$P(-1.645 < Z < 1.645) = .90 \quad \Rightarrow \quad n = [4(1.645)]^2 =$$

Thm 7.2

Let $Y_1 \cdots Y_n$ be a r.s. from a normal dist. w/ mean $= \mu$ and std. dev. $= \sigma$. Then

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n)}$$

Pf: Just restating Thm 6.4 for a r.s.

Thm If $X_1 \cdots X_k$ are ind. RVs and $Y_i = u_i(X_i)$ for $i = 1$ to k, then $Y_1 \cdots Y_k$ are independent.

Pf: (we will prove this result for X_i continuous, and $Y_i = u_i(X_i)$ being 1-1 transformations)

$$f(x_1, \dots, x_k) = \prod_{i=1}^{n} f_i(x_i) \quad \text{(by independence)}$$

Find inverse fcts: $x_1 = w_1(y_1) \cdots x_k = w_k(y_k)$

$$J = \det\begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_k} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_k} \\ \vdots & & & \\ \frac{\partial x_k}{\partial y_1} & \cdots & \cdots & \frac{\partial x_k}{\partial y_k} \end{bmatrix} = \det\begin{bmatrix} w_1'(y_1) & 0 & \cdots & 0 \\ 0 & w_2'(y_2) & 0 \cdots & 0 \\ & & & \\ 0 & \cdots & \cdots & w_k'(y_k) \end{bmatrix}$$

$$= \prod_{i=1}^{n} w_i'(y_i)$$

$$\Rightarrow g(y_1, \ldots, y_n) = f[w_1(y_1), \ldots, w_k(y_k)] \cdot \prod_{i=1}^{n} w_i'(y_i)$$

$$= \prod_{i=1}^{n} f_i[w_i(y_i)] \cdot \prod_{i=1}^{n} w_i'(y_i)$$

$$= \prod_{i=1}^{n} f_i[w_i(y_i)] \cdot w_i'(y_i)$$

$$= \prod_{i=1}^{n} h(y_i)$$

$\Rightarrow$ By factorization theorem, $Y_1 \cdots Y_k$ are independent.

<u>note:</u> We will use this result to prove $\bar{y}$ and s^2 are independent when sampling from a normal dist.

<u>Thm 7.3</u>

Let $Y_1 \cdots Y_n$ be a r.s. from $Y \sim N(\mu, \sigma^2)$ then

1. $\bar{Y}$ and S^2 are independent
2. $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \sim \chi^2_{(n-1)}$

Pf:

$$\text{① } f(y_1,\cdots,y_n) \overset{Y_i \text{ ind.}}{=} \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$ add / subtract $\bar{y}$

$$\sum_{i=1}^{n}\left[(y_i-\bar{y})-(\bar{y}-\mu)\right]^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2 - 2\sum_{i=1}^{n}(y_i-\bar{y})\cdot(\bar{y}-\mu) + \sum_{i=1}^{n}(\bar{y}-\mu)^2$$

$$= \sum(y_i-\bar{y})^2 - 2(\bar{y}-\mu)\sum(y_i-\bar{y}) + n(\bar{y}-\mu)^2$$

$$= \sum(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2$$

$$\Rightarrow f(y_1,\dots,y_n) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{1}{2\sigma^2}\left[\sum(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\right]}$$

let $u_1 = \bar{Y}$ Inverse fcts: $Y_1 = u_1 - (u_2 + \cdots + u_n)$

$u_2 = Y_2 - \bar{Y}$ $Y_2 = u_2 - u_1$

$u_3 = Y_3 - \bar{Y}$ $\vdots$

$\vdots$ $Y_n = u_n - u_1$

$u_n = Y_n - \bar{Y}$ from here, it is easy to show $J = n$

note: $(y_1-\bar{y}) + (y_2-\bar{y}) + \cdots + (y_n-\bar{y}) = 0$

$$\Rightarrow (y_1-\bar{y}) = -\sum_{i=2}^{n} u_i$$

$$\Rightarrow g(u_1,\dots,u_n) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\cdot e^{-\frac{1}{2\sigma^2}\left[\underbrace{\overbrace{\left(-\sum_{i=2}^{n}u_i\right)^2 + \sum_{i=2}^{n}u_i^2}^{\sum(y_i-\bar{y})^2}}_{k(u_2,\dots,u_n)} + \underbrace{\overbrace{n(u_1-\mu)^2}^{n(\bar{y}-\mu)^2}}_{h(u_1)}\right]}\cdot n$$

$\Rightarrow u_1$ is independent of $u_2 \cdots u_n$

$\Rightarrow \bar{y}$ " " " $(y_2-\bar{y}) \cdots (y_n-\bar{y})$

Since $y_1-\bar{y} = -\sum_{i=2}^{n}(y_i-\bar{y})$ is a fct. of $(y_2-\bar{y}) \cdots (y_n-\bar{y})$

$\Rightarrow \bar{y}$ is also independent of $y_1 - \bar{y}$

Since $s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$ is a fct of $(y_1 - \bar{y}) \cdots (y_n - \bar{y})$

$\Rightarrow \bar{y}$ is ind of s^2 ∎

② Let $W = \sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} \left(\frac{(y_i - \bar{y}) - (\bar{y} - \mu)}{\sigma} \right)^2$

$$W = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\sigma^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2}$$

$$W = \frac{(n-1)s^2}{\sigma^2} + \frac{(\bar{y} - \mu)^2}{\sigma^2/n} \rightarrow z^2 = \left(\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \right)^2$$

$$W = T + S$$

from Thm 7.2, $W \sim \chi^2_{(n)}$, $S \sim \chi^2_{(1)}$

also, T is a fct of s^2, S is a fct of $\bar{x}$, so T and S are ind.

$$M_W(t) = \frac{1}{(1-2t)^{\frac{n-1}{2}}}$$

$$M_W(t) = E[e^{tW}] = E[e^{t(T+S)}] = M_T(t) \cdot M_S(t)$$

$$\Rightarrow \frac{1}{(1-2t)^{n/2}} = M_T(t) \cdot \frac{1}{(1-2t)^{1/2}}$$

$$\Rightarrow M_T(t) = \frac{1}{(1-2T)^{n-1/2}} \Rightarrow T = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$ ∎

Example 7.5

Y = ounces of fill from a bottling machine

$Y \sim N(\mu, \sigma^2 = 1)$

Take a r.s. of $n=10$. Find b_1 and b_2 so that $P(b_1 \leq s^2 \leq b_2) = .90$

$$P\left(\frac{9 \cdot b_1}{1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \frac{9 \cdot b_2}{1}\right) = .9$$

$$P(9b_1 \leq \chi^2_{(9)} \leq 9b_2) = .9$$

$\Rightarrow 9b_1 = 3.325, \quad 9b_2 = 16.92$

$\boxed{b_1 = .369 \qquad b_2 = 1.880}$ i.e. If s^2 falls outside this interval we might believe $\sigma^2 \neq 1$

7.20

a) If $U \sim \chi^2_{(v)}$ find $E(U)$ and $V(U)$

$E(U) = v \quad var(U) = 2v$

b) If $Y_1 \cdots Y_n$ is a r.s. from $Y \sim N(\mu, \sigma)$

find $E(s^2)$, $var(s^2)$

Let $T = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)} \Rightarrow E(T) = n-1, \ var(T) = 2(n-1)$

$\Rightarrow s^2 = \frac{T \cdot \sigma^2}{(n-1)}$

$E(s^2) = \frac{\sigma^2}{(n-1)} E(T) = \frac{\sigma^2 (n-1)}{(n-1)} = \sigma^2$ (s^2 unbiased!)

$$var(s^2) = \left(\frac{\sigma^2}{(n-1)}\right)^2 var(T)$$

$$= \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

Def 7.2 Let $Z \sim N(0,1)$ and let $W \sim \chi^2_{(v)}$. Then if Z and W are independent,

$$T = \frac{Z}{\sqrt{W/v}} \sim T_{(v)}$$

2050: let $X_1 \cdots X_n$ be a r.s. from $X \sim N(\mu, \sigma^2)$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad W = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

note: Z and W are ind. because $\bar{X}$ and s^2 are ind. when sampling from a normal dist.

$$t = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T_{(n-1)}$$

Example 7.6

Y = tensile strength $Y \sim N(\mu, \sigma^2)$

for $n=6$, find the prob. $\bar{Y}$ is within $2.5/\sqrt{n}$ of μ

$$P\left(|\bar{Y}-\mu| < \frac{2S}{\sqrt{n}}\right) = P\left(\frac{-2S}{\sqrt{n}} < \bar{Y}-\mu < \frac{2S}{\sqrt{n}}\right) =$$

$P\left(-2 < \frac{\bar{y}-\mu}{s/\sqrt{n}} < 2\right) = P(-2 < T_{(5)} < 2) \approx .90$

note: $P(-2 < Z < 2) \approx .95$ i.e. the T dist. is wider and flatter than Z

7.98 Show that $T = \frac{Z}{\sqrt{W/v}}$ has density fct.

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma(v/2)}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}, \quad -\infty < t < \infty$$

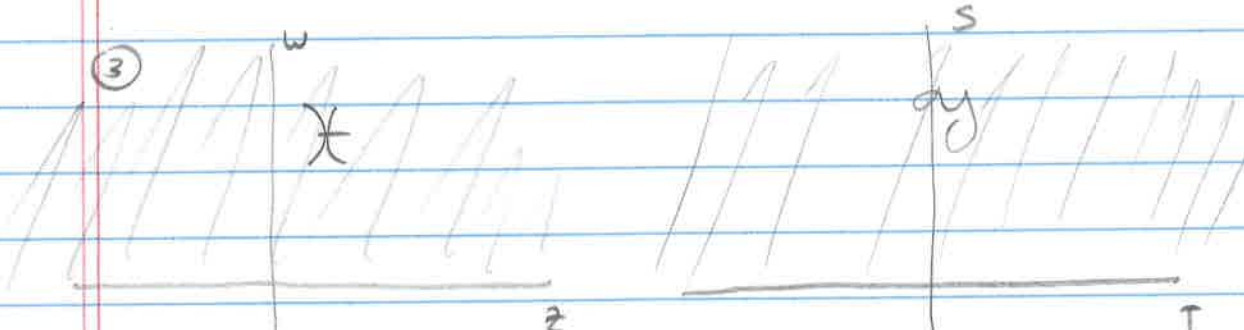
pf:

① $T = \frac{Z}{\sqrt{W/v}}$, let $S = W$

② Inverse fcts:

$$Z = T\sqrt{W/v} = T\sqrt{S/v}, \quad W = S$$

③



④ $f(z,w) = f_1(z)\cdot f_2(w)$ (because z, w are ind.)

$$f(z,w) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\right)\left(\frac{1}{\Gamma(v/2)2^{v/2}} w^{v/2-1} e^{-w/2}\right), \quad (z,w) \in \mathcal{X}$$

$$J = \det\begin{bmatrix} \sqrt{s/v} & T \cdot \frac{1}{2}\left(s/v\right)^{-1/2} \cdot \frac{1}{v} \\ 0 & 1 \end{bmatrix} = \sqrt{s/v} > 0$$

$$\Rightarrow g(s,t) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t^2 \cdot s/v)}\right)\left(\frac{1}{\Gamma(v/2)2^{v/2}} s^{v/2-1} e^{-s/2}\right)\sqrt{s/v}$$

$$\Rightarrow g_2(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t^2}{v}\right)\cdot s} \cdot \frac{1}{\Gamma(v/2)2^{v/2}} \cdot \frac{s^{v/2-1/2}}{v^{1/2}} e^{-s/2}\, ds$$

$$= \frac{1}{\sqrt{2\pi}\,\Gamma(v/2)2^{v/2}v^{1/2}} \cdot \int_0^\infty \underbrace{s^{\frac{v+1}{2}-1} \cdot e^{-s\left(\frac{t^2}{2v}+\frac{1}{2}\right)}\, ds}_{\text{think } S \sim GAM\left(\alpha = \frac{v+1}{2},\ \beta = \frac{2v}{t^2+v}\right)}$$

$$= \frac{1}{\sqrt{2\pi}\,\Gamma(v/2)2^{v/2}v^{1/2}} \cdot \Gamma\left(\frac{v+1}{2}\right)\left(\frac{2v}{t^2+v}\right)^{\frac{v+1}{2}} \cdot \cancel{\int_0^\infty \frac{1}{\Gamma\left(\frac{v+1}{2}\right)\left(\frac{2v}{t^2+v}\right)^{(v+1)/2}} \cdot s^{\frac{v+1}{2}-1} \cdot e^{-s/\left(\frac{2v}{t^2+v}\right)}\, ds}^{\,1}$$

$$= \frac{\Gamma\left(\frac{v+1}{2}\right)\cdot \cancel{2^{\frac{v+1}{2}}}\left(\frac{v}{t^2+v}\right)^{\frac{v+1}{2}}}{\cancel{2^{1/2}}\sqrt{\pi}\,\cancel{2^{v/2}}\cdot\Gamma(v/2)\cdot v^{1/2}} = \boxed{\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi \cdot v}\,\Gamma(v/2)} \cdot \left(1+\frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}, \quad -\infty < t < \infty}$$

<u>Def 7.3</u> Let $W_1 \sim \chi^2_{(v_1)}$ and $W_2 \sim \chi^2_{(v_2)}$ and W_1 and W_2 be independent. Then

$$F = \frac{W_1/v_1}{W_2/v_2} \sim F_{(v_1, v_2)}$$

<u>note:</u> Let $x_1 \cdots x_{n_1}$ be a r.s. from $X \sim N(\mu_x, \sigma_x^2)$

Let $y_1 \cdots y_{n_2}$ be a r.s. from $Y \sim N(\mu_y, \sigma_y^2)$

$$W_1 = \frac{(n_1-1)S_x^2}{\sigma_x^2} \sim \chi^2_{(n_1-1)}, \quad W_2 = \frac{(n_2-1)S_y^2}{\sigma_y^2} \sim \chi^2_{(n_2-2)}$$

W_1, W_2 are ind. because these were random samples from different pops.

$$\Rightarrow \frac{W_1/(n_1-1)}{W_2/(n_2-1)} \sim F_{(n_1-1, n_2-1)}$$

$$\Rightarrow \frac{\frac{(n_1-1)S_x^2}{\sigma_x^2} \Big/ (n_1-1)}{\frac{(n_2-1)S_y^2}{\sigma_y^2} \Big/ (n_2-1)} = \frac{S_x^2}{\sigma_x^2} \cdot \frac{\sigma_y^2}{S_y^2} \sim F_{(n_1-1, n_2-1)}$$

If $H_0: \sigma_x = \sigma_y$ vs. $H_1: \sigma_x \overset{\neq}{\underset{>}{<}} \sigma_y$

$$\Rightarrow \frac{S_x^2}{S_y^2} \sim F_{(n_1-1, n_2-1)}$$ (test statistic from 2050)

Example 7.7

$n_1 = 6 \quad n_2 = 10 \quad (\sigma_1^2 = \sigma_2^2)$

Find b so that

$$P\left(\frac{S_1^2}{S_2^2} \le b\right) = .95$$

$$= P(F_{5,9} \le b) = .95$$

$\hookrightarrow$ 3.48 (Table 7)

Show that the F dist. has density fct

$$\boxed{f(f) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma(v_1/2)\cdot\Gamma(v_2/2)} \cdot \left(\frac{v_1}{v_2}\right)^{v_1/2} f^{\,v_1/2 - 1} \left(1 + \frac{v_1}{v_2} f\right)^{-\left(\frac{v_1+v_2}{2}\right)}, \quad f > 0}$$

Outline:

1. $F = \dfrac{W_1/v_1}{W_2/v_2}$ let $G = W_2$

2. Find $h(f,g)$ using a jacobian transformation with the joint pdf of W_1, W_2

3. Find $f(f)$ by integrating $\displaystyle\int_{\text{all } g} h(f,g)$

Hint: use gamma dist

e.g. Let $y_1 \cdots y_{20}$ be a r.s. from $Y \sim N(\mu = 80, \sigma^2 = 100)$

a) Find $P(75 \le \bar{x} \le 82)$

$$P\left(\frac{75-80}{10/\sqrt{20}} \le \frac{\bar{x}-80}{10/\sqrt{20}} \le \frac{82-80}{10/\sqrt{20}}\right)$$

$$= P(-2.23 \le Z \le .89) = 1-(.0129+.1894) = .7977$$

b) Find $P\left(1980 \le \sum_{i=1}^{20}(y_i-80)^2 \le 2100\right)$

$$= P\left(\frac{1980}{100} \le \sum_{i=1}^{20}\left(\frac{y_i-80}{10}\right)^2 \le \frac{2100}{100}\right)$$

$$= P(19.8 \le \chi^2_{(20)} \le 21) = .0734$$

c) Find $P\left(1980 \le \sum_{i=1}^{20}(y_i-\bar{y})^2 \le 2100\right)$

$$P\left(\frac{1980}{100} \le \frac{19\,s^2}{\sigma^2} \le \frac{2100}{100}\right)$$

$$P(19.8 \le \chi^2_{(19)} \le 21) = .0699$$

7.3 Central Limit Thm.

Simulations in R

Thm 7.4

Let $Y_1 \ldots Y_n$ be iid with $E(Y_i) = \mu$, $V(Y_i) = \sigma^2 < \infty$.

Define

$$U_n = \frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}, \quad \bar{Y} = \frac{\sum Y_i}{n}$$

Then the d.f. of U_n converges to the standard normal d.f. as $n \to \infty$. That is,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad \text{for all } u.$$

note: The important result is that sums of iid RVs approach a normal dist. as $n \uparrow$. In particular, when we conduct inference regarding μ, the sample mean $\bar{x}$ (a fct of the sum of iid RVs) will follow a normal dist., regardless of the dist. we are actually sampling from. This makes it very easy to run hypothesis tests and compute CIs for μ.

7.42 Y = fracture strength $\mu = 14$ $\sigma = 2$

a) for $n = 100$ find

$$P(\bar{Y} > 14.5)$$

$$= P\left(\frac{\bar{Y} - 14}{2/\sqrt{100}} > \frac{14.5 - 14}{2/\sqrt{100}}\right) = P(Z > 2.5) \approx .0062$$

b) $P(b_1 < \bar{y} < b_2) = .95$

$$P\left(\frac{b_1 - 14}{.2} < \frac{\bar{y} - 14}{.2} < \frac{b_2 - 14}{.2}\right) = .95$$

$$P\left(\frac{b_1 - 14}{.2} < z < \frac{b_2 - 14}{.2}\right) = .95$$

$$P(-1.96 < z < 1.96) = .95$$

$\Rightarrow \dfrac{b_1 - 14}{.2} = -1.96$, $\dfrac{b_2 - 14}{.2} = 1.96$

$b_1 = 14 - 1.96(.2)$ $\quad$ $b_2 = 14 + 1.96(.2)$

$= 13.608$ $\quad$ $= 14.392$

$\Rightarrow P(13.608 < \bar{y} < 14.392) = .95$

i.e. we are 95% confident that $\bar{y}$ will be within

$1.96\,(2)/\sqrt{100} = .392$ of μ.

7.59 $\quad n = 50 \quad \sigma_A = \sigma_B = 2$

$P(\bar{X}_A - \bar{X}_B) \geq 1$

$$\bar{X}_A - \bar{X}_B \sim N\left(\mu = 0,\ \sigma^2 = \frac{4}{50} + \frac{4}{50} = \frac{8}{50} = \frac{4}{25}\right)$$

$$P(\bar{X}_A - \bar{X}_B > 1) = P\left(z > \frac{1-0}{2/5}\right) = P(z > 2.5) = .0062$$

e.g. FDAL for insect filth in PB : $\mu = 3$ fragments per 10 g

Y = no. of fragments in 10 g sample

If in compliance, $Y \sim POI(\lambda = 3)$

If 50 10g samples produce $\bar{y} = 3.6$, what do we conclude?

$\bar{Y} \sim N(\mu = 3, \sigma = \sqrt{3}/\sqrt{50})$ note: we don't have to "assume" σ. We know $\sigma = \sqrt{\mu}$

$$P(\bar{Y} > 3.6) = P\left(Z > \frac{3.6 - 3}{\sqrt{3}/\sqrt{50}}\right) = P(Z > 2.45) = .0071$$

i.e. $P(\bar{Y} > 3.6 \mid \mu = 3) = .0071$

$\Rightarrow$ conclude $\mu > 3$ i.e. out of compliance

7.4 Proof of CLT

Thm 7.4

Let $Y_1 \cdots Y_n$ be iid w/ $E(Y_i) = \mu$, $V(Y_i) = \sigma^2 < \infty$

Define:

$$U_n = \frac{\sum_{i=1}^{n} Y_i - n \cdot \mu}{\sigma \cdot \sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

Then the d.f. of U_n converges to the standard normal d.f.

as $n \to \infty$. That is,

$$\lim_{n \to \infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \quad \forall u$$

Pf:

Assume $m_Y(t)$ exists for $-h < t < h$

Let $W = Y - \mu \Rightarrow m_W(t) = E[e^{tW}] = E[e^{t(Y-\mu)}] = e^{-\mu t} \cdot m_Y(t)$

$m_W(0) = 1$

$m_W'(0) = E(W) = E[Y - \mu] = 0$

$m_W''(0) = E(W^2) = E[(Y-\mu)^2] = \sigma^2$

Write $m_W(t)$ as a MacClaurin Series

$$m_W(t) = \sum_{k=0}^{\infty} \frac{m_W^{(k)}(0)\,(t-0)^k}{k!} = m_W(0) + m_W'(0) \cdot t + \underbrace{\frac{m''(\xi)}{2!} t^2}_{\text{remainder term}} \qquad 0 < |\xi| < t$$

$$\Rightarrow m_W(t) = 1 + \frac{m_W''(\xi) \cdot t^2}{2}, \quad 0 < |\xi| < t$$

trick: add/subtract $\frac{\sigma^2 t^2}{2}$ ...

$$m_w(t) = 1 + \frac{\sigma^2 t^2}{2} - \frac{\sigma^2 t^2}{2} + \frac{m_w''(\varepsilon)\, t^2}{2}$$

$$= 1 + \frac{\sigma^2 t^2}{2} + \left(m_w''(\varepsilon) - \sigma^2\right) \cdot \frac{t^2}{2}$$

Now consider,

$$m_{u_n}(t) = E\left[e^{t u_n}\right] = E\left[e^{t \cdot \frac{\Sigma Y_i - n\mu}{\sqrt{n}\,\sigma}}\right] \overset{Y_i \text{ iid}}{=} \prod_{i=1}^{n} E\left[e^{t \cdot \frac{(Y_i - \mu)}{\sqrt{n}\,\sigma}}\right]$$

$$= \prod_{i=1}^{n} m_w\left(\frac{t}{\sqrt{n}\,\sigma}\right) = \left[m_w\left(t/\sqrt{n}\,\sigma\right)\right]^n, \quad -h < \frac{t}{\sqrt{n}\,\sigma} < h$$

$$= \left[1 + \frac{\sigma^2 \left(\frac{t}{\sqrt{n}\,\sigma}\right)^2}{2} + \frac{\left(m_w''(\varepsilon) - \sigma^2\right)\left(t/\sqrt{n}\,\sigma\right)^2}{2}\right]$$

$$, \quad -h < t/\sqrt{n}\,\sigma < h \quad \text{and} \quad 0 < |\varepsilon| < \left|\frac{t}{\sqrt{n}\,\sigma}\right|$$

$$= \left[1 + \frac{t^2}{2n} + \left(m_w''(\varepsilon) - \sigma^2\right) \cdot \frac{t^2}{2n\sigma^2}\right]^n, \quad -h < \frac{t}{\sqrt{n}\,\sigma} < h$$

$$\text{and } 0 < |\varepsilon| < \left|\frac{t}{\sqrt{n}\,\sigma}\right|$$

Result from calculus:

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

note: another proof w/ Taylor series

For us, $b = t^2/2$, $\quad d(n) = \left[m_w''(\varepsilon) - \sigma^2\right] \frac{t^2}{2\sigma^2}$, $\quad c = 1$

$$\lim_{n \to \infty} d(n) = \left[m_w''(0) - \sigma^2\right] \frac{t^2}{2\sigma^2} \quad \text{because } 0 < |\varepsilon| < \frac{|t|}{\sqrt{n}\,\sigma}$$

(as $n \to \infty$, $\varepsilon \to 0$)

$$\Rightarrow \lim_{n\to\infty} \phi(n) = [\sigma^2 - \sigma^2] \cdot t^2 / 2\sigma^2 = 0$$

$$\Rightarrow \lim_{n\to\infty} \left[1 + \frac{t^2/2}{n} + \frac{(m_w''(\varepsilon) - \sigma^2)\, t^2}{2\sigma^2} \Big/ n \right]^{1 \cdot n}$$

$$= e^{\frac{t^2}{2} \cdot 1} = e^{t^2/2}$$

$$\Rightarrow \lim_{n\to\infty} m_{u_n}(t) = e^{t^2/2} = m_Z(t) \; , Z \sim N(0,1)$$

$$\Rightarrow \lim_{n\to\infty} F_{u_n}(u) = \Phi(u) \quad \blacksquare$$

Show:

$$\lim_{n\to\infty}\left(1+b/n+\frac{\phi(n)}{n}\right)^{cn}=e^{bc} \quad \text{if } \lim_{n\to\infty}\phi(n)=0$$

Pf:

$$\lim_{n\to\infty}\left(1+b/n+\frac{\phi(n)}{n}\right)^{cn}=1^{\infty} \quad (\text{indeterminate form})$$

$$\lim_{n\to\infty}\left(1+b/n+\frac{\phi(n)}{n}\right)^{cn}=\lim_{n\to\infty}e^{\ln\left(1+b/n+\frac{\phi(n)}{n}\right)^{cn}}$$

$$=e^{\lim_{n\to\infty} cn\cdot\ln\left(1+b/n+\frac{\phi(n)}{n}\right)}$$

Now, consider the Maclaurin series of $\ln(1+x)$

$f(x)=\ln(1+x) \qquad f(0)=0$

$f'(x)=\frac{1}{1+x} \qquad f'(0)=1 \qquad (0!)$

$f''(x)=\frac{-1}{(1+x)^2} \qquad f''(0)=-1 \quad (-1!)$

$f^{(3)}(x)=\frac{2}{(1+x)^3} \qquad f^{(3)}(0)=2 \quad (2!)$

$f^{(4)}(x)=\frac{-3!}{(1+x)^4} \qquad f^{(4)}(0)=-3!$

$\vdots$

$f^{(k)}(x)=\frac{(-1)^{k+1}(k-1)!}{(1+x)^k} \qquad f^{(k)}(0)=(-1)^{k+1}(k-1)!$

2

$$\ln(1+x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)\,(x-0)^k}{k!}$$

$$= 0 + \frac{1\cdot x}{1!} + \frac{-1\cdot x^2}{2!} + \frac{2!\,x^3}{3!} + \frac{-3!\,x^4}{4} + \cdots$$

$$= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

$$= \sum_{n=1}^{\infty} \frac{x^n}{n}(-1)^{n+1}$$

$$\Rightarrow \ln\left(1+\frac{b}{n}+\frac{\phi(n)}{n}\right)$$

$$= \left(\frac{b}{n}+\frac{\phi(n)}{n}\right) - \frac{1}{2}\left(\frac{b}{n}+\frac{\phi(n)}{n}\right)^2 + \frac{1}{3}\left(\frac{b}{n}+\frac{\phi(n)}{n}\right)^3 - \frac{1}{4}\left(\frac{b}{n}+\frac{\phi(n)}{n}\right)^4 + \cdots$$

$$\Rightarrow cn\ln\left(1+\frac{b}{n}+\frac{\phi(n)}{n}\right)$$

$$= c(b+\phi(n)) - \frac{1}{2}\cdot\frac{1}{n}(b+\phi(n))^2 + \frac{1}{3n}(b+\phi(n))^3 - \frac{1}{4n}(b+\phi(n))^4 + \cdots$$

$$\Rightarrow \lim_{n\to\infty} cn\cdot\ln\left(1+b/n+\frac{\phi(n)}{n}\right) = cb$$

$$\Rightarrow e^{\lim_{n\to\infty} cn\cdot\ln(1+b/n+\phi(n)/n)} = e^{cb}$$

①

7.5 The Normal approx. to the Binomial

Let $X_1, \ldots, X_n$ be a r.s. from $X_i \sim Bin(1,p)$ $\quad E(X_i)=p$, $Var(X_i)=p(1-p)$

Let $Y = \sum_{i=1}^{n} X_i \Rightarrow Y \sim Bin(n,p)$

However, since Y is the sum iid RVs, CLT says that for large n ..

- $\dfrac{Y - n\cdot p}{\sqrt{np(1-p)}}$ is approx $N(0,1)$ i.e. Y is approx. normal $(\mu = np,\ \sigma^2 = np(1-p))$

- also $\bar{Y} = \hat{p}$ is approx. normal $\left(\mu = p,\ \sigma^2 = \dfrac{p(1-p)}{n}\right)$

How large does n need to be? Different books give different criterion.

① $np \ge 5$ and $n(1-p) \ge 5$

② $np(1-p) \ge 10$

③ Our book:

$$0 < p \pm 3\cdot\sqrt{\frac{p(1-p)}{n}} < 1$$

or equivalently ... $\boxed{n > 9\left(\dfrac{\max(p,q)}{\min(p,q)}\right)}$ see 7.70

note: The closer p gets to .5 and the larger n gets, the better the approx.

$\boxed{\text{Simulations}}$ $\quad n > 9\left(\dfrac{.95}{.05}\right) = 171$

$$n > 9\left(\frac{.5}{.5}\right) = 9$$

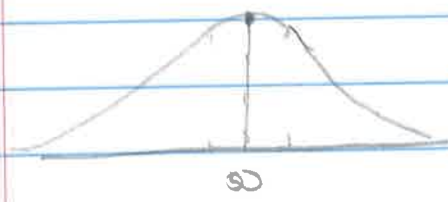
②

Continuity Correction

Spse $Y \sim Bin(n=100, p=.5)$

$\mu = 50 \quad \sigma = 5 \qquad n > 9\left(\frac{.5}{.5}\right) = 9$ ✓

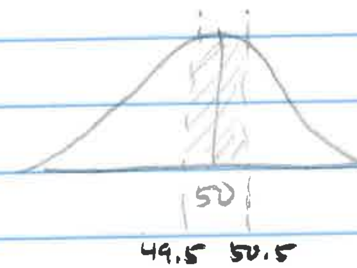
$\Rightarrow$ Y is approx. normal, $\mu = 50$, $\sigma = 5$



To approx. $P(Y=50)$ using a normal dist. we get 0

To get a prob. using a cont. dist., we need to create an interval.

Use $P(49.5 < Y < 50.5)$



e.g. (2050)

American Airlines claims 90% of all flights from Orl. to LA are on-time.

If we sample 150 flights, approx. the prob. that fewer than 125 are on-time.

Y = no. on-time flights $\quad Y \sim Bin(n=150, p=.9)$

$9\left(\frac{.9}{.1}\right) = 81 \qquad n > 81$ ✓ $\Rightarrow$ use normal approx.

$Y \sim N(\mu = 135, \sigma = 3.67)$

$P(Y \leq 124) \approx P(Y < 124.5) =$

$P\left(Z < \frac{124.5 - 135}{3.67}\right) =$

$P(Z < -2.86) = .0021$ $\boxed{\text{Conclude? } p < .9}$

note: exact prob. using R/excel is .0039. The approx. was off by only .0018, but the relative error was

$$\frac{.0018}{.0039} = 46\% \quad \text{not too good}$$

$\Rightarrow$ For p near 0 or 1, we need larger values of n.

e.g.

Casino has roulette table

- bet on odd $P(\text{win}) = 18/38$. This is an "even" bet, i.e. the payoff is 1:1

- For 1000 plays, find the prob. the casino loses money

Let Y = no. of odd, $Y \sim Bin(n = 1000, p = 18/38)$

$9\left(\frac{20/38}{18/38}\right) < n \Rightarrow$ use normal approx. $\mu = 1000(18/38) = 473.68$

$$P(Y \geq 501) \approx P(Y > 500.5) \approx P\left(Z > \frac{500.5 - 473.68}{\underbrace{\sqrt{1000(18/38)(20/38)}}_{15.8}}\right)$$

$= P(Z > 1.6986) = .0447$ Exact prob =

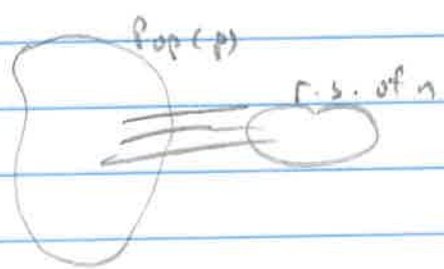
Why do we care about this? Just use R/excel and get exact prob.?

We often are interested in estimating a pop. proportion, p.

e.g. Election Polls

Defectives

Smokers, etc.



$$\hat{p} = \frac{\overbrace{x_1 + \cdots + x_n}^{\text{0s and 1s}}}{n} = \frac{Y}{n} = \bar{Y}$$

$\hat{p} = \bar{Y}$ is approx. normal, $\mu = p$, $\sigma = \sqrt{\frac{p(1-p)}{n}}$

(for large n)

Use this result to construct CIs for p:

$$\boxed{\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$ using $\hat{p}$ for p

note: when sampling from large pops., we are approx. the hypergeometric with the binomial. If N is small ($n > .05N$) then use

$$\underbrace{\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{FPCF}}}_{\text{SD for HYP}}$$