Ding: E52-303
OH: T 11:30-12:30

Graham: OH: W 3-4

Sample statistics

$\underbrace{T(x_1, \ldots, x_n)}_{\text{outcome}}$, where $X_1 = x_1, \ldots, X_n = x_n$

$\underbrace{T(X_1, \ldots, X_n)}_{\text{random variable}}$

Interested in characterizing the properties of $T(X_1, \ldots, X_n)$

1] Exact distribution theory. (small sample results)

2] approximate distribution for $T(X_1, \ldots, X_n)$
- asymptotics

Exact sampling results

Defn 5.1.1: The random variables $X_1, \ldots, X_n$ are called a random sample of size n from population $f_X(x)$ if $X_1, \ldots, X_n$ are mutually independent with the same marginal distributions. Referred to as iid with pdf $f_X(x)$.

Remarks: 1] Simplest possible sampling scheme

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

2] Economics - this assumption is difficult to make.
time series - $f(x_2|x_1) = f(x_2)$ unlikely

Sums of random variables

Lemma 5.2.5: Let $X_1, \dots, X_n$ be a r.s. Let $g(x)$ be such that mean and variance exist. Then

$$E\left[\sum_{i=1}^{n} g(X_i)\right] = n\,E[g(X_i)] \quad \text{any } i$$

$$Var\left[\sum_{i=1}^{n} g(X_i)\right] = n\,Var(g(X_i)) \quad \text{any } i$$

To see those:

$$E\left[\sum_{i=1}^{n} g(X_i)\right] = \sum_{i=1}^{n} \underbrace{E[g(X_i)]}_{\substack{\text{equal for all } i \\ \text{by identicality}}} = n\,E[g(X_i)]$$

$$Var\left(\sum_{i=1}^{n} g(X_i)\right) = E\left[\left(\sum_{i=1}^{n} g(X_i) - E\left[\sum_{i=1}^{n} g(X_i)\right]\right)^2\right]$$

$$= E\left[\left\{\sum_{i=1}^{n} \left(g(X_i) - E[g(X_i)]\right)\right\}^2\right]$$

$$= E\left[\left\{\sum_{i=1}^{n} \left(g(X_i) - E[g(X_i)]\right)\right\}^2\right]$$

$$+ \sum_{j \neq i} \underbrace{E\left[\left(g(X_i) - E[g(X_i)]\right)\left(g(X_j) - E[g(X_j)]\right)\right]}_{=0 \text{ by independence.}}$$

$$= \sum_{i=1}^{n} E\left[\left(g(X_i) - E[g(X_i)]\right)^2\right]$$

$$= \sum_{i=1}^{n} Var(g(X_i)) = n\,Var(g(X_i))$$

by identicality

e.g. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. $E[\bar{X}] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n} n E[X_i]$

$$= E[X_i] = \mu$$

$$Var(\bar{X}) = n\,Var\left(\frac{X_i}{n}\right) = \frac{1}{n}Var(X_i) = \frac{\sigma^2}{n}$$

Sampling distribution for $\bar{X}$

Thm 5.2.7: Let $X_1, \dots, X_n$ be rs with mgf $M_X(t)$.

Then $M_{\bar{X}}(t) = [M_X(\frac{t}{n})]^n$

Pf: $M_{\bar{X}}(t) = E[\exp\{t\bar{X}\}] = E[\exp\{\frac{t}{n}\sum_{i=1}^{n} X_i\}]$

$$= E\left[\prod_{i=1}^{n}\exp\{\tfrac{t}{n}X_i\}\right] = \prod_{i=1}^{n} E[\exp\{\tfrac{t}{n}X_i\}]$$

$$= \prod_{i=1}^{n} M_{X_i}(\tfrac{t}{n}) = [M_X(\tfrac{t}{n})]^n \quad \square$$

e.g. $X_i \overset{iid}{\sim} N(0,1)$

$$M_{X_i} = \exp\{\tfrac{t^2}{2}\}$$

$$M_{\bar{X}} = \left[e^{(\frac{t}{n})^2/2}\right]^n = \left[e^{\frac{t^2}{2n^2}}\right]^n = e^{\frac{t^2}{2n}} = e^{(\frac{t}{\sqrt{n}})^2/2}$$

which is the MGF for $Y \sim N(0, \frac{1}{n})$

e.g. $X \sim Cauchy(0,1)$

$$\varphi_X(t) = e^{-|t|}$$

$\bar{X}_2 = \frac{1}{2}(X_1 + X_2)$, $X_i \overset{iid}{\sim} Cauchy(0,1)$

$$\varphi_{\bar{X}_2}(t) = E\left[e^{i\frac{1}{2}(X_1+X_2)t}\right] = E\left[\exp\left\{i\frac{X_1}{2}t\right\}\right]E\left[\exp\left\{i\frac{X_2}{2}t\right\}\right]$$

$$= \left[\varphi_X\left(\frac{t}{2}\right)\right]^2 = e^{-\left|\frac{t}{2}\right|}e^{-\left|\frac{t}{2}\right|} = e^{-|t|}$$

Thus $\bar{X}_n \sim$ Cauchy $(0,1)$.

e.g. Sampling from a normal distribution with unknown variance.

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$T(X) = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

<u>2 steps</u>

1. Joint distribution of $\bar{X} - \mu$ and S.
2. Transform to distribution $T(X)$.

<u>Thm 5.3.1</u> $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then

a) $\bar{X}, s^2$ are independent

b) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

c) $(n-1)\frac{s^2}{\sigma^2} \sim \chi^2(n-1)$

$$T(X) = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\frac{s}{\sigma}} \Rightarrow T = \frac{U}{\sqrt{V/p}}$$

$$\Rightarrow U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}, \quad \frac{S}{\sigma} = \sqrt{V/p} \Rightarrow V = p\frac{S^2}{\sigma^2}, \quad p = n-1$$

$$\Rightarrow T = \frac{U}{\sqrt{V/p}}, \quad \text{Let } W = V$$

inverse functions: $v = w$ $\quad 0 < v < +\infty$

$u = t\sqrt{w/p}$ $\quad -\infty < u < +\infty$

joint distribution

$$f_{U,V}(u,v) = f_U(u) f_V(v)$$
$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\} \cdot \frac{1}{\Gamma(\frac{p}{2})2^{p/2}} v^{(\frac{p}{2}-1)} \exp\left\{-\frac{v}{2}\right\}$$

$$J = \begin{vmatrix} \frac{\partial u}{\partial t} & \frac{\partial u}{\partial w} \\ \frac{\partial v}{\partial t} & \frac{\partial v}{\partial w} \end{vmatrix} = \begin{vmatrix} \sqrt{w/p} & \frac{t}{2\sqrt{wp}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{w}{p}}$$

$$f_{T,W}(t,w) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}t^2\frac{w}{p}\right\}\frac{1}{\Gamma(\frac{p}{2})2^{p/2}} w^{(\frac{p}{2}-1)} e^{-\left\{\frac{w}{2}\right\}} \sqrt{\frac{w}{p}}$$

$$f_T(t) = \int_0^\infty f_{T,W}(t,w)\,dw$$

$$= \frac{1}{\sqrt{2\pi}\,\Gamma(\frac{p}{2})2^{\frac{p}{2}}p^{1/2}}\int_0^\infty \exp\left\{-\frac{t^2}{2}\frac{w}{p} - \frac{w}{2}\right\} w^{(\frac{p}{2}-1+\frac{1}{2})}\,dw$$

$$= \frac{1}{\sqrt{2\pi}\,\Gamma(\frac{p}{2})2^{\frac{p}{2}}p^{1/2}}\int_0^\infty \exp\left\{-\frac{1}{2}\left(1+\frac{t^2}{p}\right)w\right\} w^{(\frac{p+1}{2}-1)}\,dw$$

Set $\alpha = \frac{p+1}{2}$, $\frac{1}{\beta} = \frac{1}{2}\left(1 + \frac{t^2}{p}\right)$

$$\Rightarrow f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)\left(\frac{1}{1+t^2/p}\right)^{\left(\frac{p+1}{2}\right)}}{\Gamma\left(\frac{p}{2}\right)\sqrt{p\pi}} = \text{student's } t \text{ distribution.}$$

Ullah - covers all the tricks of the trade for those small sample transformations.