Ch8: Estimation

§8.1 an estimator

We seek an unknown population parameter θ (eg, μ, σ^2, p).
The rule used to approx or guess θ is called an estimator.
Using said rule on set of observations is called the estimate.

ex: Want $\theta = \mu$.
estimator: $\bar{X} = \frac{1}{n} \sum X_i$.

an estimate: Given 7 random observations of X, $\{Y_1, \ldots, Y_7\}$, $\bar{Y} = \frac{1}{7} \sum_1^7 Y_i$ is an estimate.

Rmk: Last example is a point estimator.

There are also interval estimators.
eg. Confidence intervals
$\theta = \mu$.
$$\bar{X} - Sz_{*} \le \mu \le \bar{X} + Sz_{*}$$

§8.2 trying to understand if ~~an~~ point estimator is any good

"good"

Let $\hat{\theta}$ be a point estimator for a parameter θ.
(e.g.: $\theta = \mu$, $\hat{\theta} = \overline{X}$)

① One goal of a good estimator is that $E[\hat{\theta}] = \theta$.
(e.g. by CLT, $E[\overline{X}] = \mu$)
But this is not always the case.

def: If $E(\hat{\theta}) = \theta$, we say $\hat{\theta}$ is an <u>unbiased</u> point estimator. (e.g. $\overline{X}$ unbiased)

If $\hat{\theta}$ is a biased point estimator, we define the <u>bias</u> to be $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

② another goal of a good estimator might be that its "spread" of observation is tightly packed about θ. (Talking about variance).

def: The <u>mean square error</u> of $\hat{\theta}$ is $MSE(\hat{\theta}) = E[(\hat{\theta}-\theta)^2]$

Cor: $MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$.

Pf: $MSE(\hat{\theta}) = E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2)$
$= E(\hat{\theta}^2) - 2E(\hat{\theta})\theta + \theta^2$
by linearity and fact, since θ is constant,
$E(\theta) = \theta$

$$= E(\hat{\theta}^2) - (E(\hat{\theta}))^2 + (E(\hat{\theta}))^2 - 2E(\hat{\theta})\theta + \theta^2$$

$$= \underbrace{E(\hat{\theta}^2) - (E(\hat{\theta}))^2}_{V(\hat{\theta})} + \underbrace{(E(\hat{\theta}))^2 - 2E(\hat{\theta})\theta + \theta^2}_{(E(\hat{\theta}) - \theta)^2} \quad \blacksquare$$

Rmk: we have decomposed MSE by

$$MSE(\hat{\theta}) = \underbrace{V(\hat{\theta})}_{\text{"precision"}} + \underbrace{B(\hat{\theta})^2}_{\text{"accuracy"}}$$

Idea of a "best" estimator is tricky.
We'd like MSE to be as small as possible.
Usually this is an impossible problem because MSE depends on the unknown θ.

ex: (population proportions)
Want p
To estimate, let $Y = \begin{cases} 1 & \text{"success"} \\ 0 & \text{"failure"} \end{cases}$

and $\{Y_i\}$ an iid random sample.

estimator 1: $\hat{p} = \frac{1}{n} \sum_1^n Y_i$

Recall $\sum Y_i$ is binom(n,p) and that
$E(\sum Y_i) = np$ and $V(\sum Y_i) = npq$
$= np(1-p)$.

Bias? $E(\hat{p}) = E\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n} E(\sum Y_i)$

$= \frac{1}{n} \cdot np = p$

$E(\hat{p}) = p$!. $\hat{p}$ is unbiased!

For MSE need $V(\hat{p})$.

$$V(\hat{p}) = V\left(\frac{1}{n}\sum Y_i\right) = \left(\frac{1}{n}\right)^2 V\left(\sum Y_i\right)$$

$$= \frac{1}{n^2} \, n p(1-p) = \frac{p(1-p)}{n},$$

and

$$MSE[\hat{p}] = V(\hat{p}) + B(\hat{p})^2$$

$$= \frac{p(1-p)}{n} + 0 = \frac{p(1-p)}{n}$$

estimator 2: $\tilde{p} = \dfrac{\sum Y_i + 1}{n+2}$

Bias? $E[\tilde{p}] = \dfrac{E[\sum Y_i] + E[1]}{n+2}$

$$= \frac{np+1}{n+2}$$

$\neq p$ hence biased and

$$B(\tilde{p}) = E(\tilde{p}) - p$$

$$= \frac{np+1}{n+2} - p = \frac{np+1-p(n+2)}{n+2}$$

$$= \frac{1-2p}{n+2}.$$

For MSE,

$$V(\tilde{p}) = V\left(\frac{\sum Y_i + 1}{n+2}\right) = V\left(\frac{\sum Y_i}{n+2}\right)$$

$$= \frac{np(1-p)}{(n+2)^2}$$

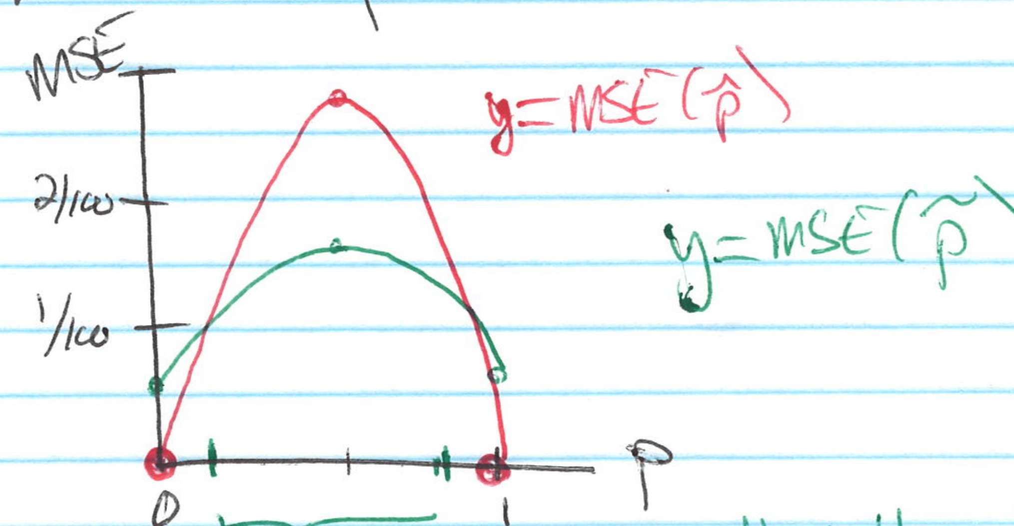
δpζ

and $MSE(\hat{p}) = \dfrac{np(1-p)}{(n+2)^2} + \left(\dfrac{1-2p}{n+2}\right)^2$

$= \dfrac{1 + (n-4)p + (4-n)p^2}{(n+2)^2}$

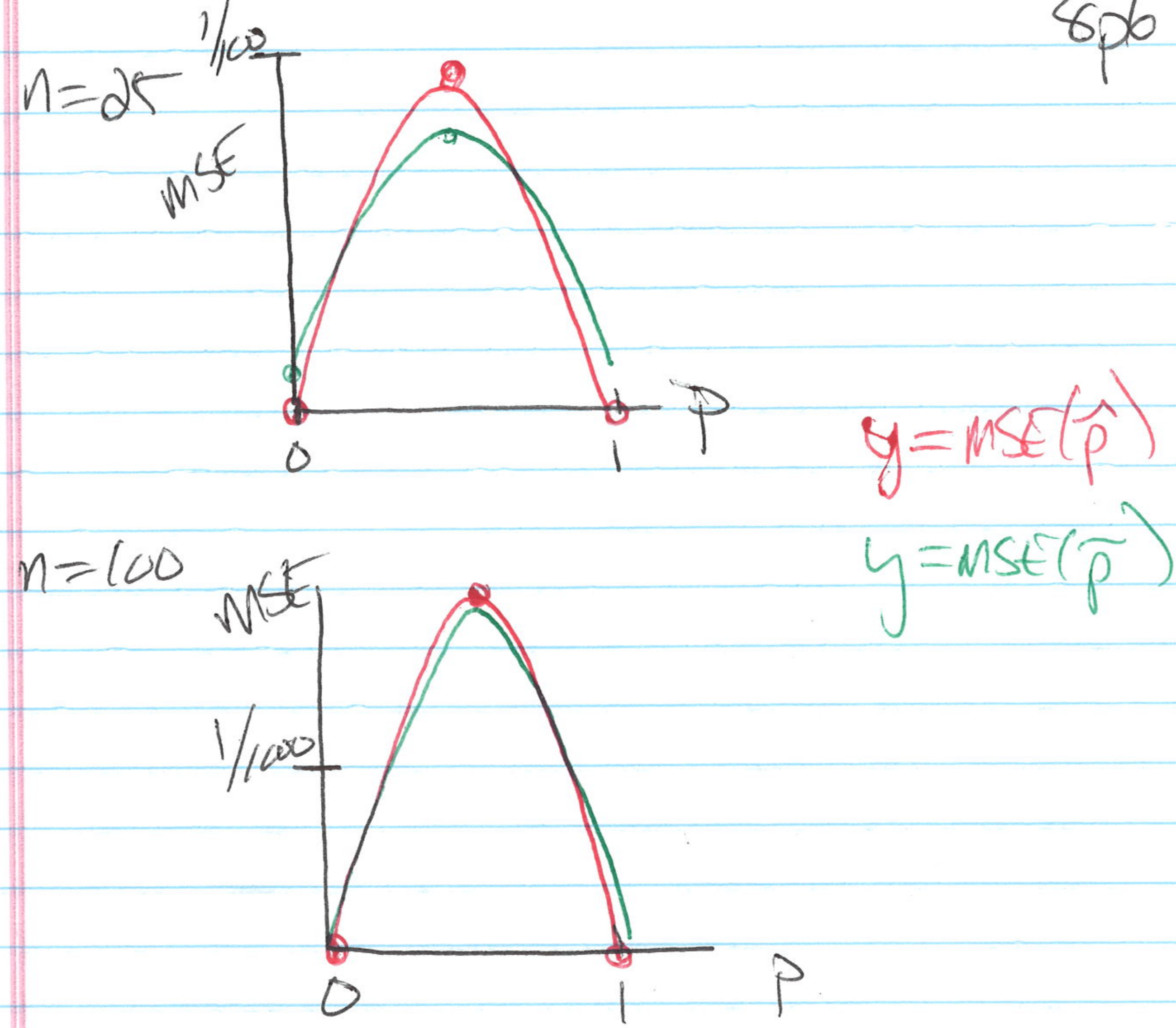
Which estimator is better?
Note for both, $MSE(\hat{\theta})(n,p)$, are fcns of n and p.
Let $y = MSE(\tilde{p})(n,p)$ and $y = MSE(\hat{p})(n,p)$.
Note when $n>4$, the graphs of both are downward parabolas w/ vertex at $p = 1/2$.

Fix $n=10$ and plot.



in this zone $\tilde{p}$ is the better estimate
~ out here $\hat{p}$ is better.

* for small n, and p near 50%, $\hat{p}$ is better



* larger n, becoming indistinguishable.

§ 8.3

Table 8.1 pg 397 : Common unbiased estimators for μ, p, $\mu_1 - \mu_2$, $p_1 - p_2$.

ex. Variance and the $n-1$.

Natural def'n is $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$.

Why do we use $n-1$? Check $B(S^2)$. Is $E(S^2) = \sigma^2$?

Let $E(X) = \mu$ and $V(X) = \sigma^2$ and $X_1, \dots, X_n$ iid random sample.

$$E(S^2) = E\left(\frac{1}{n}\sum (X_i - \bar{X})^2\right) = \frac{1}{n} \underbrace{E\left(\sum (X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right)}_{\text{focus on this } *}$$

$$* = E\left(\sum X_i^2 - 2\sum X_i \bar{X} + \sum \bar{X}^2\right)$$
$$= \underset{①}{E\left(\sum X_i^2\right)} - 2\underset{②}{E\left(\sum X_i \bar{X}\right)} + \underset{③}{E\left(\sum \bar{X}^2\right)}$$

① and ③ use same trick $V(Y) = E(Y^2) - [E(Y)]^2$.

① $E\left(\sum X_i^2\right) = \sum E(X_i^2) = \sum \left(V(X_i) + (E(X_i))^2\right)$
$= \sum (\sigma^2 + \mu^2) = n(\sigma^2 + \mu^2)$

③ $E\left(\sum \bar{X}^2\right) = \sum \left(V(\bar{X}) + E(\bar{X})^2\right)$
$= n\left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + n\mu^2$

② $-2E\left(\sum x_i \left(\frac{1}{n}\sum_j X_j\right)\right)$

$= -\frac{2}{n}E\left(\sum_i \sum_j X_i X_j\right)$ n^2 terms in $\sum\sum$.

$= -\frac{2}{n}E\left(\underbrace{\sum_i X_i^2}_{n \text{ terms}} + \underbrace{\sum\sum_{i \neq j} X_i X_j}_{n^2-n \text{ terms}}\right)$

but X_i's iid
$\Rightarrow Cov(X_i, X_j) = 0$
or $E(X_i X_j) = E(X_i)E(X_j)$

$= -\frac{2}{n}\left(\sum_i E(X_i)^2 + \sum\sum_{i \neq j} E(X_i X_j)\right)$

$= -\frac{2}{n}\left(n\underbrace{(V(X_i) + \mu^2)}_{\text{like in ①, ③}} + (n^2-n)\mu\mu\right.$.

$= -2(\sigma^2 + \mu^2) - 2(n-1)\mu^2$

So $\cancel{E(\ )}$ $*$ $= \underline{n\sigma^2} + n\mu^2 \underline{-2\sigma^2} - 2\mu^2 - 2n\mu^2 + 2\mu^2 \underline{+\sigma^2} + n\mu^2$

$= (n-1)\sigma^2$

Then $E(S^2) = \cancel{\ }\frac{1}{n} \cdot * = \left(\frac{n-1}{n}\right)\sigma^2$

$\neq \sigma^2$ ... Biased.

To make an unbiased estimator, rescall the ~~sum~~ summation of the natural def'n.

$S^2 = \frac{1}{n-1}\sum_i^n (X_i - \overline{X})^2$.

§ 8.4 Probability statements about $\hat{\theta}$

def: Given any $\hat{\theta}$ the exact (theoretical) error is defined $\epsilon = |\hat{\theta} - \theta|$.

Note ϵ is itself a r.v. and we can make probability statements about it.

Obviously want ϵ small. $|\hat{\theta} - \theta| < b$.
Consider $P(|\hat{\theta} - \theta| < b)$

Chebyshev $P(|\hat{\theta} - \theta| < k\sigma) > 1 - \frac{1}{k^2}$

$k = 2$ a common choice in practice.

$$P(|\hat{\theta} - \theta| < 2\sigma) > 0.75$$

In practice 2σ is usually much better than this (See Table 8.2) when underlying dist'n of $\hat{\theta}$ is symmetric.
Rmk: we call σ the <u>standard error</u>.

Also in practice, we use $s^2 \approx \sigma^2$ (more later)

ex: (difference in means)
2 iid random samples.
$n_1 = 100$, $\bar{Y}_1 = 26400$, $S_1^2 = 1,440,000$

$n_2 = 200$, $\bar{Y}_2 = 25,100$, $S_2^2 = 1,960,000$

Estimate the difference of means $\mu_1 - \mu_2$.

Here $\bar{Y}_1 - \bar{Y}_2 = 26400 - 25100 = 1300$ a point estimate.

Recall $\sigma^2_{\bar{Y}_1 - \bar{Y}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

So $\sigma^2 \approx S^2_{\bar{Y}_1 - \bar{Y}_2} = \frac{1,400,00}{100} + \frac{1,960,000}{200}$

$= 22800$

So $S_{\bar{Y}_1 - \bar{Y}_2} \approx 151$.

an interval estimate for $\mu_1 - \mu_2$

$1300 \pm 2(151) = 1300 \pm 302$

or $(998, 1602)$

§ 8.5 Confidence Intervals

Want $P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1-\alpha$

$\hat{\theta}_L, \hat{\theta}_U$, ~~upper~~ lower and upper confidence limits.
$1-\alpha$ the confidence coef.

When we know how $\hat{\theta}$ is distributed, we can use standardization methods to find the limits.

disc: the pivotal method for finding confidence intervals.

1. We know how a r.v. Y is distributed but not some underlying parameter θ.

2. we can convert to a prob. dist'n that does not depend on θ.

ex: $\bar{X}$ distributed $N(\mu, \sigma^2/n)$

via $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $\xrightarrow{\text{convert}}$ $N(0,1)$
indep of μ

ex: $Y_1, \ldots, Y_n$ a random sample of size n from a uniform dist'n on the interval $(0, \theta)$.

Want an estimate for θ.
Use $\hat{\theta} = \max(Y_1, \ldots, Y_n)$

8p12.

We know how $\hat{\theta}$ is distributed, but clearly dependent upon θ.

$$f(y) = \frac{1}{\theta} \Rightarrow F(y) = P(Y \le y) = \int_0^y \frac{1}{\theta} = \frac{y}{\theta},$$
$$0 \le y \le \theta.$$

The max order stat $\hat{\theta}$ has CDF

$$P(\hat{\theta} \le \omega) = P(Y_1 \le \omega, Y_2 \le \omega, \ldots, Y_n \le \omega) = P(Y_1 \le \omega)^n$$
$$= \begin{cases} 0, & \omega < 0 \\ (\omega/\theta)^n, & 0 \le \omega \le \theta \\ 1, & \omega > \theta. \end{cases}$$

Use a change of variable to find an associated pivotal distribution: $U = \frac{\hat{\theta}}{\theta}$

$$P(U \le u) = P(\hat{\theta}/\theta \le u)$$
$$= P(\hat{\theta} \le \theta u) = \left(\frac{\theta u}{\theta}\right)^n = u^n \text{ on } 0 \le u \le 1$$
$$= \begin{cases} 0, & u < 0 \\ u^n, & 0 \le u \le 1 \\ 1, & u > 1. \end{cases}$$

* pivotal CDF of U, no longer depends on θ

We use U's CDF to construct a confidence interval.

Find a 95% lower confidence interval for θ.

Want $P(\underbrace{\hat{\Theta}_L \le \theta}) = 0.95$

an example of a one-sided confidence interval

Using U's CDF, $P(U \le u) = 0.95$
when $u^n = 0.95$ or $u = (0.95)^{1/n}$.

So $P\left(\frac{\hat{\Theta}}{\theta} \le (0.95)^{1/n}\right) = 0.95$

equivalently, $P\left(\frac{\hat{\Theta}}{(0.95)^{1/n}} \le \theta\right) = 0.95$

Our confidence interval: $\frac{Y_{(n)}}{(0.95)^{1/n}} \le \theta$.

eg. (concrete)
For random sample 0.76, 0.88, 1.68, 1.74, 1.78

A 95% lower C.I. for θ is given by
$n = 5$, $Y_{(5)} = 1.78$

$\Rightarrow \frac{1.78}{(0.95)^{1/5}} \le \theta$ or $1.798 \le \theta$

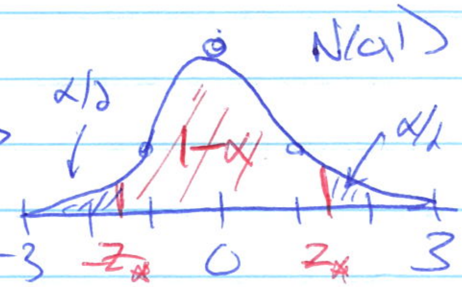
§ 8.6 Large Sample C. I.

The unbiased point estimators for μ, p, $\mu_1 - \mu_2$, $p_1 - p_2$ all have near Normal distributions by the CLT.

Moreover, using $Z = \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$, Z is a pivotal quantity

For 2-sided C.I.,

$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha \iff$



Some common standard errors:

90% C.I. $\Rightarrow z_{0.05} = 1.645$

95% C.I. $\Rightarrow z_{0.025} = 1.960$

99% C.I. $\Rightarrow z_{0.005} = 2.576$

Then $-z_{\alpha/2} \le \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \le z_{\alpha/2}$

$$\Longrightarrow \underbrace{\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}}_{\hat{\theta}_L} \le \theta \le \underbrace{\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}}_{\hat{\theta}_U}$$

Of course, we don't know $\sigma_{\hat\theta}$ exactly.
For large samples we use $\sigma_{\hat\theta} \approx S_{\hat\theta}$.

ex: $\bar{x} = 19.07$, $S^2 = 10.60$ with $n = 32$.
Recall $\bar{x}$ distributed by $N(\mu, \sigma^2/n)$.
Here $\approx$ distributed by $N(\mu, S^2/n) = N\left(\mu, \frac{10.60}{32}\right)$.

For a 95% C.I. for μ, $\sigma_{\bar{x}} \approx \sqrt{\frac{10.60}{32}} \approx 0.576$

$$\bar{x} \pm z_* \sigma_{\bar{x}} \longleftrightarrow 19.07 \pm (1.96)(0.576)$$
$$19.07 \pm 1.128$$
or $(17.94, 20.20)$

ex: differences in population proportions.

Estimating $p_1 - p_2$ by $\hat{p}_1 - \hat{p}_2$ from samples of size n_1 and n_2 respectively.

$$\hat\theta \pm z_{\alpha/2}\, \sigma_{\hat\theta} \Longrightarrow \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2}$$

$$\Longrightarrow \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\underbrace{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}_{\text{depends on the unknowns } p_1, p_2.}}$$

Two standard fixes:

(1) $pq = p(1-p) \leq 1/4$. easy to show by calculus.

Yields a "max" error via $\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}$.

However, in practice we use following as smaller confidence intervals are desirable.

(2) When n_i is large enough, we can use $\hat{p}_i$ for p_i

The $1-\alpha$ C.I. is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

§8.7 - Sample Size and Error Bounds.

For our C.I., we have $\hat{\theta} \pm z_{\alpha/2}\, \sigma_{\hat{\theta}}$.

- $\sigma_{\hat{\theta}}$ is the standard error
- $E = z_{\alpha/2}\, \sigma_{\hat{\theta}}$ is the error bounds of our C.I.

In real-world, we choose our confidence level $1-\alpha$ and/or the error E ahead of time.
This leads to having to choose an appropriate sample size n.

ex: for $\bar{X}$ ... dist'n is $N(\mu, \sigma^2/n)$.

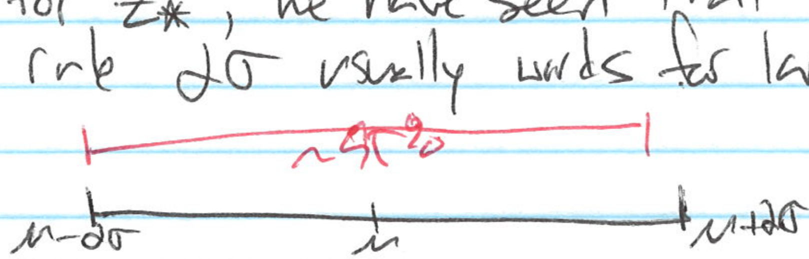
$$\sigma_{\bar{X}} = \sqrt{\sigma^2/n} \quad \text{and} \quad E = z_* \sqrt{\sigma^2/n}$$

$$\longleftrightarrow \quad n = \frac{z_* \sigma^2}{E^2}.$$

For desired E, need to choose $n \geq \left\lceil \frac{z_* \sigma^2}{E^2} \right\rceil$

another issue: we don't really know $\sigma_{\hat{\theta}}^2$ and maybe even z_*.

1. for z_*, we have seen that the empirical rule 2σ usually works for large n.



(2) For $\sigma_{\hat{\theta}}$, either

a) use old data or a "pilot" sample s^2

or

b) Use 1/4 the spread of the data set

i.e. $\sigma_{\hat{\theta}} \approx \frac{1}{4}(\max(Y_i) - \min(Y_i))$

However, for proportions we can do even better.

ex: proportions.

For $\hat{p} = \frac{\sum Y_i}{n}$ (relative frequency)

$\hat{p}$ distributed $N(p, pq/n)$.

$E = z_* \sqrt{\frac{pq}{n}} \longleftrightarrow n \geq \left\lceil \frac{z_*^2 p(1-p)}{E^2} \right\rceil$

Again, since $p(1-p) \leq 1/4$ on $[0,1]$, we choose $n \geq \frac{z_*^2}{4E}$.

eg: public opinion polls $(\pm 3\%)$.

For 95% C.I., $z_* = 1.960$ and $n \geq \frac{(1.96)^2}{4(0.03)^2} \approx 1068$

For 99%, $z_* = 2.576$ and $n \geq 1843$.

§ 8.9 Small Sample C.I. for means (μ and $\mu_1 - \mu_2$).

So far for $\bar{X}$ when n is large we have used

$$\frac{\hat{\theta}-\mu}{\sigma/\sqrt{n}} \approx \frac{\hat{\theta}-\mu}{S/\sqrt{n}}$$

↳ for small n, no longer precise enough

But we set ourselves up for this issue in Ch 7.
Recall we know

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \text{ is distributed } N(0,1)$$

and $V = \dfrac{(n-1)S^2}{\sigma^2}$ is distributed χ^2 w/ $n-1$ d.f.

and Z and V are indy.

$$\text{Let } T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{1}{S/\sigma}$$

$$= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot \frac{1}{\sqrt{S^2/\sigma^2}}$$

$$= Z \cdot \frac{1}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}}$$

We write $T = Z \Big/ \sqrt{V/r}$ where $r = n-1$.

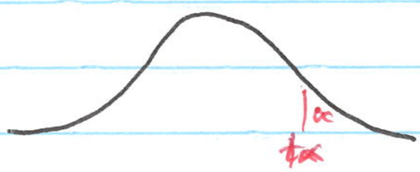
def: We say T has a t-sampling dist'n (t-dist'n) w/ $n-1$ degrees of freedom and its

8 p22

pdf is given by
$$f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma(r/2)}\left(1+\frac{t^2}{r}\right)^{-(r+1)/2}$$

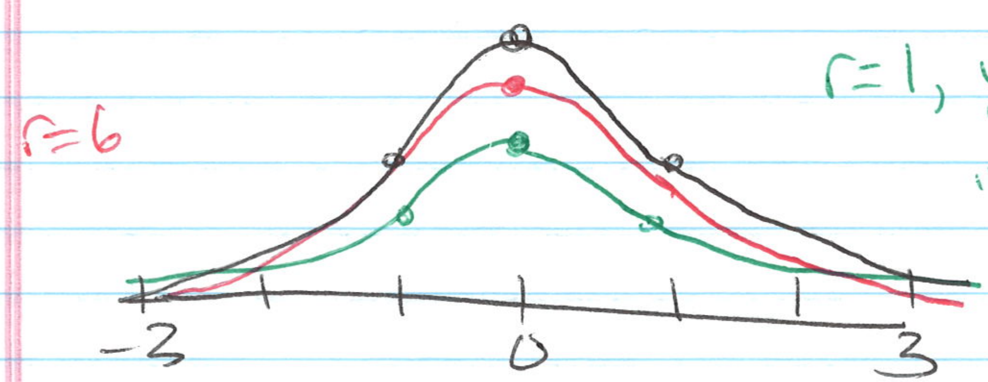
Rmks: ① We won't work w/ $f_T(t)$ explicitly.
Note this dist'n is pivotal ... indep of μ and σ^2.
We pick our $t_{\alpha/2}$ from Table 5, p 849
To do so need α and $d.f. = r = n-1$.

eg.



df	$t_{.1}$	$t_{.05}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
10	1.372	1.812	2.228	2.764	3.169
⋮ table only goes to $30 = \infty$					
z-scores! → 30	1.282	1.645	1.960	2.326	2.576

② t-dist'n look like "fat tailed" Normal Curves.



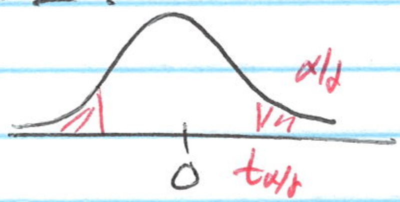
$r=1$, $y = \dfrac{1}{\sqrt{\pi}\,(1+t^2)}$
"Cauchy dist'n"

FACT: as $r \to \infty$, T-dist'n $\longrightarrow N(0,1)$.

8,221

Fun fact: Cauchy dist'n is "famous"
$r=1$, $\mu_T = 0$ but $\sigma_T^2 = \infty$.

disc. Small n C.I.

For 2-sided, 

, $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}}$

$$P(-t_{\alpha/2} \le T \le t_{\alpha/2}) = 1-\alpha$$

$$\Longrightarrow \quad \bar{X} \pm t_{\alpha/2}\left(S/\sqrt{n}\right).$$

ex: $n=10$, $\bar{X} = 3.22$, $S = 1.17$.

Find a 95% C.I for μ.
Table 5: 95% $\rightarrow t_{0.025}$ w/ $df = 10-1 = 9$.
use $t_{0.025} = 2.262$.
and then

$$\bar{X} \pm t_{*}\frac{S}{\sqrt{n}} \Longrightarrow 3.22 \pm (2.262)\left(\frac{1.17}{\sqrt{10}}\right)$$

$$3.22 \pm 0.84$$
or $(2.38, 4.06)$

8pd2

BONUS MATH: Where from ... the t-dist'n.

For the math curious, it comes down to a fancy change of variables.

Recall from last semester, if X, Y are indy then the joint pdf is $f(x,y) = f_X(x) f_Y(y)$.
Here $T = Z(V/r)^{-1/2}$ and Z, V are indy (as $\bar{X}$ and S are by Fisher's).

To derive f_T, start w/ the joint pdf of Z and V, a χ^2 dist'n of r d.f.
Then $f(y,z) = \frac{1}{\Gamma(r/2) 2^{r/2}} y^{r/2-1} e^{-r/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

$$0 < y < \infty, \quad -\infty < z < \infty.$$

Then let $t = z(v/r)^{-1/2}$, $v = y$.

$$P(Y < y, Z < z) = \int_{-\infty}^{y} \int_{-\infty}^{z} f(y,z)\, dy\, dz$$

$$= \int_{-\infty}^{t} \int_{-\infty}^{v} f(y(t,v), z(t,v)) \begin{vmatrix} y_t & y_v \\ z_t & z_v \end{vmatrix}^{-1} dt\, dv.$$

the Jacobian from Calc III

(Recall from 420: $\int_a^b f(x)\,dx = \int_{h(a)}^{h(b)} f(h^{-1}(u)) \frac{dh^{-1}(u)}{dx}\, du$

$u = h(x)$

disc: difference in means, small sample.

Start w/ X, $E[X] = \mu_X$, $V[X] = \sigma_X^2$, Normal
and Y, $E[Y] = \mu_Y$, $V[Y] = \sigma_Y^2$, Normal.

Now we have 2 different iid random samples
$\bar{X}$ and $S_{\bar{X}}$ w/ n_1
$\bar{Y}$ and $S_{\bar{Y}}$ w/ n_2.

We have the unbiased estimator $\bar{X} - \bar{Y}$
w/ $V(\bar{X}-\bar{Y}) = \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}$.

Then $Z = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n_1 + \sigma_Y^2/n_2}}$ as a pivotal quantity.

Again, for small n_i, using $S_i \approx \sigma_i$ is not precise enough.

Assumption: Let $\sigma_1 = \sigma_2$.

Then $Z = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Now we need to construct a point estimator for σ.

def: the pooled estimator

$$S_p^2 = \frac{\sum(X_i-\bar{X})^2 + \sum(Y_i-\bar{Y})^2}{r}$$

Then $$S_p^2 = \frac{(n_1-1)S_{\bar{X}}^2 + (n_2-1)S_{\bar{Y}}^2}{r}.$$

Claim: S_p^2 is unbiased when $r = n_1+n_2-2$
$(= n_1-1 + n_2-1)$

i.e. $E[S_p^2] = \sigma^2$.

Reason: Really the same type of computation we did earlier.

or...

think about this as a product dist'n of 2 indy χ^2 dist'ns, one of n_1-1 df the other n_2-1. Then the product dist'n will have n_1-1+n_2-1 d.f.

Using $S_p^2 = \frac{(n_1-1)S_{\bar{X}}^2 + (n_2-1)S_{\bar{Y}}^2}{n_1+n_2-2}$, we can define the t-statistic for $\mu_X-\mu_Y$ as

$$T = \frac{(\bar{X}-\bar{Y}) - (\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$ which has

n_1+n_2-2 d.f.

§8.9 Confidence Intervals for σ^2

We have $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$ is unbiased

and $V = \frac{(n-1)S^2}{\sigma^2}$ is a pivotal quantity

with χ^2 dist'n of $n-1$ df.

For a C.I., we use

$$P\left(\chi_L^2 \le \frac{(n-1)S^2}{\sigma^2} \le \chi_u^2\right) = 1-\alpha.$$

As w/ t-dist'ns, we need both α and df $n-1$.
Most common construction is to pick symmetric "tails".

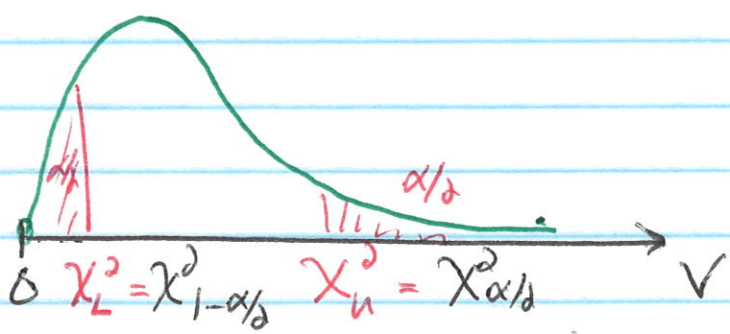


Table 6, pg 850–851 gives upper tail probs.

χ_L^2 on 850, χ_u^2 on 851

also

df: 1–30, 40, 50, ..., 100.

8 pdfs

Then $\frac{\chi_L^2}{(n-1)S^2} \le \frac{1}{\sigma^2} \le \frac{\chi_u^2}{(n-1)S^2}$

$\Rightarrow \frac{(n-1)S^2}{\chi_L^2} \ge \frac{1}{\sigma^2} \ge \frac{(n-1)S^2}{\chi_u^2}$

or $\boxed{\frac{(n-1)S^2}{\chi_u^2} \le \sigma^2 \le \frac{(n-1)S^2}{\chi_L^2}}$.

ex: Construct 90% C.I for μ and σ^2.

85.4, 86.8, 86.1, 85.3, 84.8, 86.0

$n=6$

$\bar{X} = 85.7\overline{3}$, $S^2 = 0.502\overline{6}$, $S = 0.7089$

90% $\Rightarrow \alpha = 0.10$

$df = 6-1 = 5$

Table 6, $\chi_L^2 = \chi^2_{0.95} = 1.145476$

$\chi_u^2 = \chi^2_{0.05} = 11.0705$

For σ^2, $\frac{5 \cdot (0.502\overline{6})}{11.0705} \le \sigma^2 \le \frac{5 \cdot (0.502\overline{6})}{1.145476}$

$0.227 \le \sigma^2 \le 2.194$

For μ, use t-dist'n w/ df=5

$\bar{X} \pm t_{0.05} S$, $85.\cancel{73} \pm (2.015)(0.7089)$

± 1.428