Ch 7 Review

Our 1st goal in 326 is to learn how to estimate global parameters of a "population" like μ and σ. To do this, we need to understand how the r.v. we use to estimate them are distributed.

$$\text{eg. param.} \begin{cases} \mu & , \bar{X} = \frac{1}{n}\sum X_i \\ \sigma^2 & , S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2 \end{cases} \text{stats.}$$

§7.2

① sample means

Thm 7.1: Let $Y_1, \ldots, Y_n$ be a random sample from $N(\mu, \sigma^2)$. Then $\bar{Y}$ is distributed by $N(\mu, \sigma^2/n)$.

* the dist'h of sample means $\bar{Y}$ also Normal!

Reason: Linearity of Expectation and Variance Properties

disc: Recall in working w/ $N(\mu, \sigma^2)$, we learned it was easier to standardize everything via z-scores: $Z = \frac{X-\mu}{\sigma}$

and Z distributed $N(0,1)$

② sample variance

Recall standard deviation σ is a measure of spread of the r.v. and is derived from

$(Y_i - \bar{Y})^2$ terms. In all math, we normalize to take the units out of things.

$$U_i = \frac{Y_i - \bar{Y}}{S} \approx \frac{Y_i - \mu}{\sigma} = Z_i$$

data driven
a stat!

not a stat
depends on population parameters.
But theoretically, easier to pretend we start here.

Thm 7.2: If $Y_1, \ldots, Y_n$ are a random sample from $N(\mu, \sigma^2)$ then $U = \sum Z_i^2 = \sum \left(\frac{Y_i - \mu}{\sigma}\right)^2$

has a χ^2-dist'n with n degrees of freedom.

Reason: Old HW Z_i^2 is χ^2 w/ $df = 1$.
By product of mgf, $\sum Z_i^2$ is χ^2 w/ $df = n$.

Recall: χ^2 dist'n is a Gamma$(\nu/2, 2)$
where $\nu := df$.

Now, to get to sample variance, we algebra.

$$S^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$$
$$\approx \frac{1}{n-1} \sum (Y_i - \mu)^2$$

$$\frac{S^2}{\sigma^2} \approx \frac{1}{n-1} \sum \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

$$\frac{(n-1) S^2}{\sigma^2} \approx \sum Z_i^2 \; !$$

Showing it is okay to replace $\bar{Y}$ w/ μ is the point of the proof of...

Thm 7.3 (Fisher's Thm - the dist'n of sample variance S)
If $Y_1, \ldots, Y_n$ a random sample fm $N(\mu, \sigma)$, then ① $\frac{(n-1)S^2}{\sigma^2}$ has χ^2 dist'n w/ $(n-1)$ d.f.

and ② $\bar{Y}$ and S^2 are independent r.v.

③ t-dist'n and F-dist'n.

These are defined here, but not derived.
Skipping for the moment.
Also Law of Large #s.

§ 7.4 The CLT.

The reason Normal dist'n play an outsized role in applied stats is that the dist'n of $\bar{Y}$ can be made nearly normal no matter the underlying dist'n of Y (does NOT need to start life Normal). So nearly normal, we just pretend it is.

Thm: The Central Limit Theorem
Let $Y_1, \ldots, Y_n$ be independent and identically distributed (i.i.d.) r.v. w/ $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$.
Define $U_n = \dfrac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \dfrac{\sum Y_i - n\mu}{\sigma\sqrt{n}}$.
The dist'n fn of U_n converges to ~~the~~ $N(0,1)$ as $n \to \infty$.

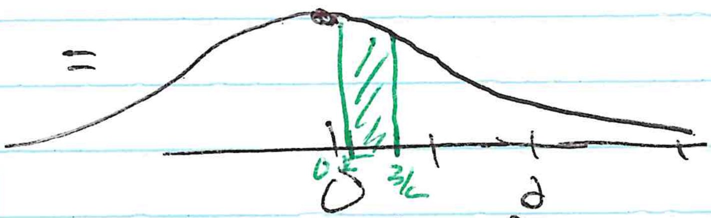
ex: Let $\bar{X}$ denote the mean of a random sample of size $n=15$ from the dist'n whose pdf is $f(x) = \frac{3}{2}x^2$, $-1 \le x \le 1$.

Can be shown $\mu = 0$ and $\sigma^2 = \frac{3}{5}$ for X.

To compute $P(0.03 \le \bar{X} \le 0.15)$ we use the CLT and assume $\bar{X}$ is distributed $N(\mu, \sigma^2/n) = N(0, 1/25)$

Then $P\left(\frac{0.03-0}{\sqrt{1/25}} \le Z \le \frac{0.15-0}{1/5}\right)$

$= P(0.15 \le Z \le 0.75)$

$=$ 

$\approx$ Table 4 (0.15) $-$ Table 4 (0.75)

(upper tails table)

$= 0.7734 - 0.5596$

$= 0.2138$

disc: about n.

Surprisingly, "large" n doesn't have to be that large.

Usually, for any r.v. X and any dist'n

$\boxed{n \ge 30}$ is enough.

That is, when $n \ge 30$ the CLT ~~yields a~~ says $N(\mu, \sigma^2/n)$ yields a good approximation of $\bar{X}$.

When X is symmetric, unimodal, and continuous an n of 4 or 5 is often enough.

§7.5: Normal Approx. to Binomial Dist'ns.