Lecture 20 Math 241 11/24/15 Recall $Z \sim N(0,1)$, $M_Z(t) = e^{t^2/2}$

$\boxed{L1}$

$X_1, \ldots, X_n \overset{iid}{\sim}$ something with finite μ, σ^2

$$C_n := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \Bigg\} \text{ the standardized } \bar{X}$$

Goal: what is the distribution of C_n as n gets big?

from last time

$$C_n = \ldots = \frac{Z_1}{\sqrt{n}} + \ldots + \frac{Z_n}{\sqrt{n}} \quad \text{s.t.} \quad Z_i := \frac{X_i - \mu}{\sigma} \Bigg\} \text{ the stdized } X \text{ r.v.}$$

$$M_{C_n}(t) = M_{\sum}(t) = \left(M_{\frac{Z}{\sqrt{n}}}(t)\right)^n = \left(M_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n =$$

$a = \frac{1}{\sqrt{n}}$

$$M_X(t) = 1 + \frac{t\,E(X)}{1!} + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \dots$$

$$\left(M_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \underbrace{\frac{t\,E(Z)}{\sqrt{n}\,1!}}_{=0} + \underbrace{\frac{t^2 E(Z^2)}{(\sqrt{n})^2\,2!}}_{=1} + \overbrace{\frac{t^3 E(Z^3)}{(\sqrt{n})^3\,3!} + \frac{t^4 E(Z^4)}{(\sqrt{n})^4\,4!} + \dots}^{e(n)}\right)^n$$

due to standard

$$= \left(1 + \frac{t^2/2}{n} + e(n)\right)^n = \left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)\right)^n$$

Note $e(n) = o\left(\frac{1}{n}\right)$ which means $\lim\limits_{n\to\infty} \dfrac{e(n)}{\frac{1}{n}} = 0$

$e(1) = 1.1$

$$\lim \frac{\frac{a}{n^{1.5}}}{\frac{1}{n}} + \frac{\frac{b}{n^2}}{\frac{1}{n}} + \dots = \frac{a}{\sqrt{n}} + \frac{b}{n} + \dots = 0 \checkmark$$

n	lim
100	5.04
1000	4.48
10^6	3.49
10^9	3.08
10^{12}	2.90 $\to$ 2.718

$\Rightarrow$ $e(n)$ goes to 0 faster than $\frac{1}{n}$

We care about $\lim_{n\to\infty} C_n$

$$\lim_{n\to\infty} M_{C_n}(t) = \lim_{n\to\infty}\left(1+\frac{t^2/2}{n}+o\left(\frac{1}{n}\right)\right)^n$$

We know $e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$

What is $\lim_{n\to\infty}\left(1+\frac{x}{n}+o\left(\frac{1}{n}\right)\right)^n = e^x$ as well but it may converge slowly!!!!!

does this grow quick enough to affect the limit?

turns out ... no so

$$\lim_{n\to\infty}\left(1+\frac{t^2/2}{n}+o\left(\frac{1}{n}\right)\right)^n = e^{t^2/2} \Rightarrow C_n \sim N(0,1)$$

as n gets large

<u>Central Limit Thm</u> crown jewel of M241!

So $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \overset{\text{approx}}{\sim} N(0,1)$ if n is large... How large? (see HW)

$$\Rightarrow Z = \overset{\sim N(0,1)}{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}} \Rightarrow \bar{X} = \frac{\sigma}{\sqrt{n}} Z + \mu \overset{approx}{\sim} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

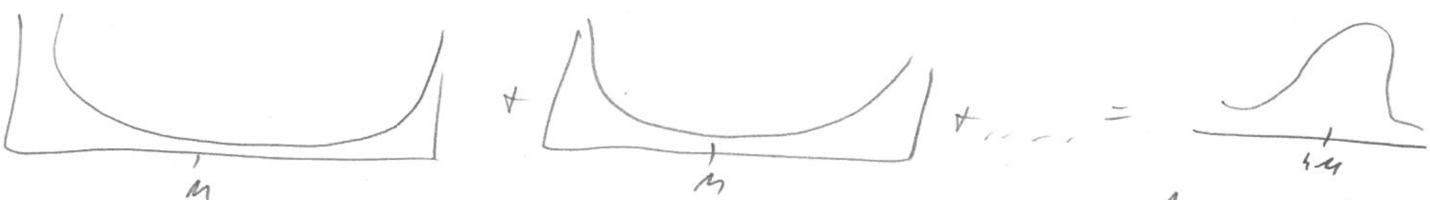
$\bar{X}$ is approx Norm distr. if n is large

$\Rightarrow$ begining of Statistics

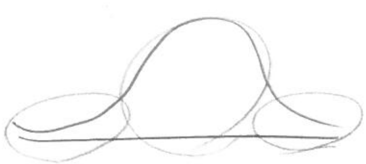
$$T_n = n\bar{X} = \sigma\sqrt{n} Z + n\mu \overset{approx}{\sim} N\left(n\mu, (\sigma\sqrt{n})^2\right)$$

T is also approx normal distr. if n is large

Imagine $X \sim$ Bathtub



Amazing!!



means extremes

$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

e^{-x^2} makes small tails

It is the normal balance between means & extremes

already it shows!

Ex Random walk 100 steps. What's the prob you're more than 10 steps away from origin? $X_1, \dots, X_{100} \sim$ Random $\mu = 0,\ \sigma = 1$ $\quad T \overset{approx}{\sim} N(n\mu, (\sigma\sqrt{n})^2) = N(0, 10^2)$

$$P(|T| > 10) = 2P(T > 10) = 2P(Z > 1) = 2 \cdot .16 = \boxed{.32}$$

Ex $X_1, \dots, X_{30} \overset{iid}{\sim} \text{Geom}\left(\frac{1}{2}\right)$

What's the prob on avg I wait more than 2.75 flips

$$\mu = \frac{1}{p} = \frac{1}{\frac{1}{2}} = 2, \quad \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-\frac{1}{2}}{\left(\frac{1}{2}\right)^2}} = \sqrt{2} = 1.414 \qquad \frac{\sigma}{\sqrt{n}} = \frac{1.414}{\sqrt{30}} = .258$$

$$\bar{X} \overset{approx}{\sim} N(2, .258) \quad P(\bar{X} > 2.75) = P\left(\frac{\bar{X} - 2}{.258} > \frac{2.75 - 2}{.258}\right) = P(Z > 3) = \boxed{.0015\%}$$

Ex $X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(p)$. What is the prob $\bar{X}$ > or < something?

$$\mu = p \quad \sigma = \sqrt{p(1-p)} \Rightarrow \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\bar{X} \overset{approx}{\sim} N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$$

e.g. Shipments are late 2% of the time. Over 10,000 orders. What is the prob more than 3% late on avg?

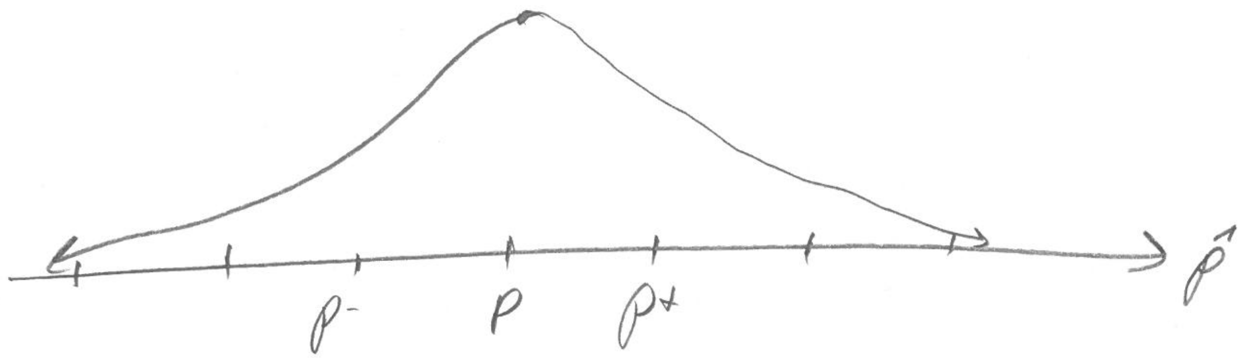
$$1.3 \times 10^{12}$$

$$\bar{X} \sim \left(.02, \left(\sqrt{\frac{.02 \cdot .98}{10{,}000}} = .0014\right)^2\right) \quad P(\bar{X} > 3\%) = P\left(Z > \frac{.03 - .02}{.0014}\right) = P(Z > 7.14) \approx 0$$

→ $\bar{X}$ is special. It's the avg # of successes.

It gets a special notation: $\hat{p}$, $\quad \hat{p} := \frac{\#1\text{'s}}{n} = \bar{X}$

Ex: Who likes mushrooms?

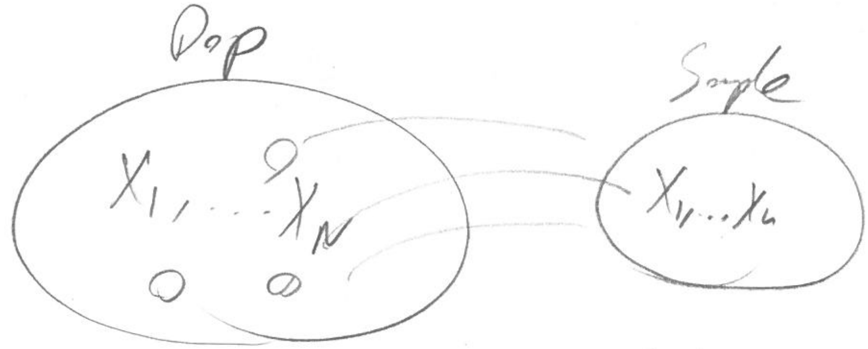


$f(\hat{p}) = ?$

$\hat{p}$ falls anywhere here...

Prob ↑

Stats ↓



$|N| = |M|$ infinite

$|n| << \infty$ ← sample size

Our focus... of more true...

↓

$X_1, \dots, X_n \overset{iid}{\sim} Bern(p)$

if n is big enough, CLT kicks in (see HW)

$\hat{P} \sim N(\dots)$

$\hat{p}$ is a random var

Representative sample?

Sampling strategy? Biased samples

Goal: Use $\hat{p}$ to infer/guess p, the parameter.

L?

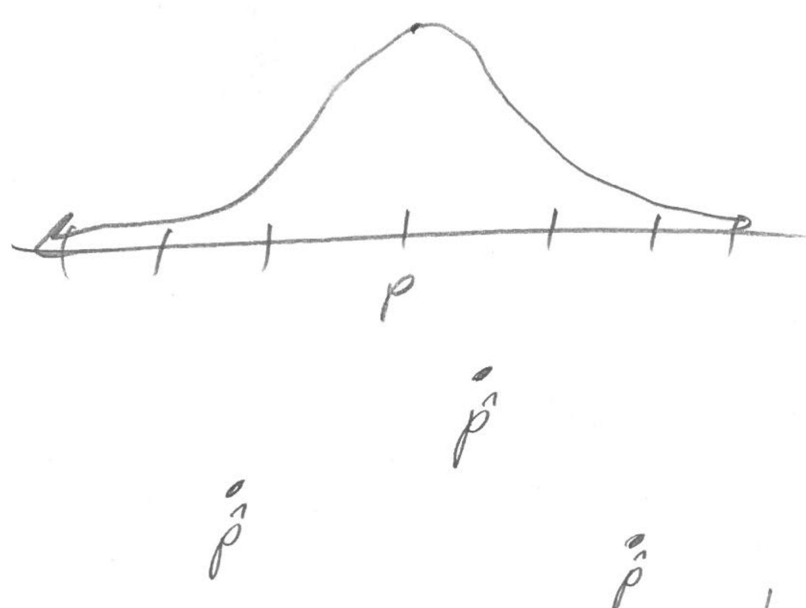
What are pop parameters?? Single values that we don't know.

Why? We don't know the pop. We can't see the whole pop.

Best guess $\hat{p} \approx p$ since $\bar{X} \approx \mu$ by LLN

"point estimate" best estimate of a single pt.

What about an "interval estimate"? An estimate for p over an interval? Imagine if I did the following:



by def. margin of error

$$\left[\hat{p} - \sqrt{\frac{p(1-p)}{n}},\ \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right] = \left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right]$$

pt est.