11/30/2017

Lecture 21.

How to use CLT

$\lim_{x \to \infty} \frac{1}{x} = 0$

$\frac{1}{x} \approx 0$ if x is "large" → based on what precision you want

EXAM: If n=2, can't use normal, will be 20's, 30's

If $X_1, \ldots, X_n \overset{iid}{\sim}$ with mean $\mu$, variance $\sigma^2$ and if n is "large" enough

(2) $\overline{X} \overset{d}{\approx} N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$

(3) $T \overset{d}{\approx} N(n\mu, (\sigma\sqrt{n})^2)$

You begin at O, the origin and you take 100 steps where each is either forward or backward w.p 1/2. What is the probability you will be more than 10 steps away after the 100 steps?

$X_1, \ldots, X_{100} \overset{iid}{\sim} \begin{cases} 1 & w.p\ 1/2 \\ -1 & w.p\ 1/2 \end{cases} \Rightarrow \mu = 0 \quad \sigma^2 = 1 \Rightarrow \sigma = 1$

$T = X_1 + X_2 + \ldots + X_{100}$

$P(T > 10) + P(T < -10) = P\left(\frac{T-0}{10} > \frac{10-0}{10}\right) + P\left(\frac{T-0}{10} < \frac{-10-0}{10}\right) = P(Z > 1) + P(Z < -$

standardized normal

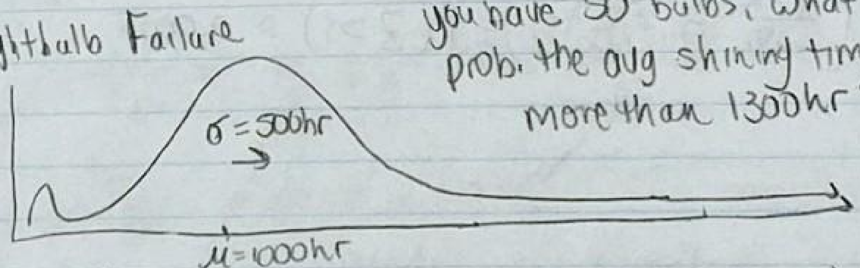
$.16 + .16$

$\boxed{= .32}$

By the CLT $T \overset{d}{\approx} N(n\mu, (\sigma\sqrt{n})^2)$

$N(100 \cdot 0, (1 \cdot \sqrt{100})^2) = N(0, 10^2)$

X: Lightbulb Failure




you have 50 bulbs. What is the prob. the avg shining time is more than 1300hr?

$X_1 \ldots X_{50} \overset{iid}{\sim}$

$P(\overline{X} > 1300) = P\left(\frac{\overline{X} - 1000}{70.7} > \frac{1300 - 1000}{70.7}\right) = P(Z > 4.2) \approx 0$

(2) $\overline{X} \overset{d}{\approx} N(\mu, (\frac{\sigma}{\sqrt{n}})^2) = N\left(1000, \left(\frac{500}{\sqrt{50}}\right)^2\right) = N(1000, 70.7^2)$

Shipments are late 2% of the time. In 10,000 orders, what is the prob. more than 3% are late?

$X_1, \ldots X_{10{,}000} \overset{iid}{\sim} Bern(0.02) \Rightarrow \mu = 0.02,\ \sigma = \sqrt{0.02(1-0.02)} = 0.14$

$$P(\bar{X} > 0.3) = P\left(\frac{\bar{X}-0.02}{.0014} > \frac{0.03-0.02}{0.0014}\right) = P(Z > 7.14) \approx 0$$

$$② \ \bar{X} \overset{d}{\approx} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) = N\left(0.02, \left(\frac{0.14}{\sqrt{10000}}\right)^2\right) = N(0.02, 0.0014^2)$$

$$\bar{X} \overset{d}{\approx} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

if X's are Bernoullis

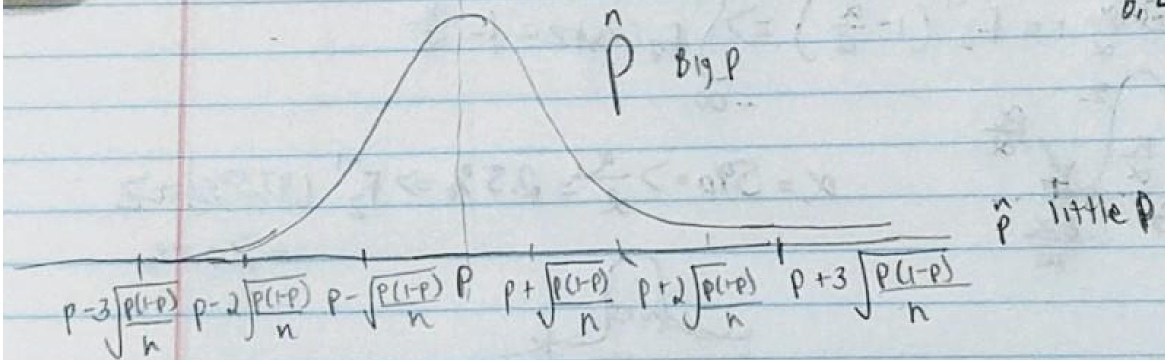
$$\hat{P} \overset{d}{\approx} N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$$

$\bar{X}$ is the avg r.v, $\bar{x}$ is a realization

$\hat{P}$ is the sample proportion r.v, $\hat{p}$ is a realization

sample proportion $\rightarrow \hat{p} \leftarrow$ estimate $= \bar{x} = \frac{1+1+0+0+0}{5}$

$0.4 \in [0,1]$

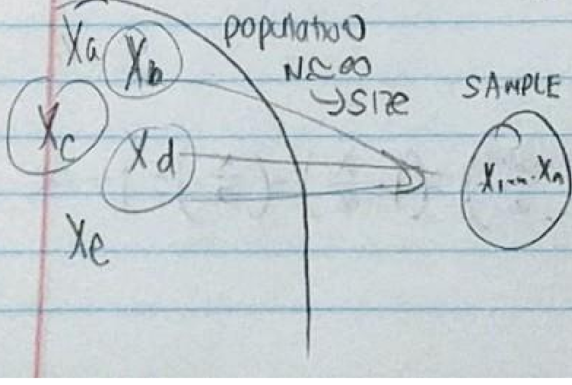



PROB ↑

STAT ↓

We cannot know p. But... we can

Statistical inference
① Make a best guess $\Rightarrow p \approx \hat{p}$ (estimate)
② Create a window (range of likely values of p
③ Test ideas about p

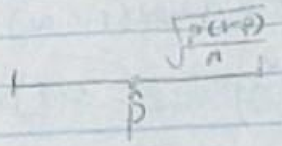



$n << N$ but large enough to use the CLT

$$\hat{p} = \frac{X_1 + \ldots X_n}{n}$$

sample must be a "simple random sample"
i.e totally at random otherwise you get bias

window ②




What is the probability this window "covers" (i.e $p \in$ window).

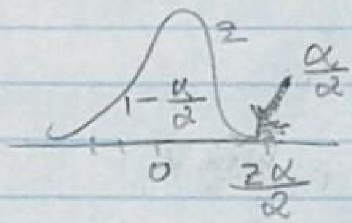
$$P\left(p \in \hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right) = P\left(\hat{p} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= P\left(-\sqrt{\frac{p(1-p)}{n}} \le p - \hat{p} \le \sqrt{\frac{p(1-p)}{n}}\right) = P\left(-1 \le \frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \le 1\right) = P(-1 \le -Z \le 1)$$

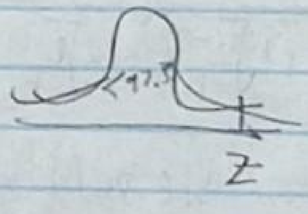
$$= P(1 \ge Z \ge -1)$$

$P\left(Z \in [-1,1]\right) = .68$

Define $Z_{\frac{\alpha}{2}} := F_Z^{-1}\left(1-\frac{\alpha}{2}\right) \Rightarrow \int_{-\infty}^{Z_{\frac{\alpha}{2}}} f_Z(z)\,dz = 1-\frac{\alpha}{2}$




$\alpha = 5\% \Rightarrow \frac{\alpha}{2} = 2.5\% \Rightarrow F_Z^{-1}(97.5\%) = Z$
$= Z_{0.5\%}$




$$= P\left(Z \in \left[-\frac{Z_\alpha}{2}, \frac{Z_\alpha}{2}\right]\right) = F_Z\left(\frac{Z_\alpha}{2}\right) - F_Z\left(\frac{-Z_\alpha}{2}\right) = \left(1-\frac{\alpha}{2}\right) - \left(\frac{\alpha}{2}\right) = 1-\alpha$$

$$\left[\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \boxed{CI_{p,1-\alpha} := \left[\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]}$$

100 yr of debate

CI = confidence interval
(2 sided, 1 proportion CI)

$[3 \pm 2] = [1, 5]$

* $CI_{p,95\%} := \left[\hat{p} \pm Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$