Nov 29 Lecture

How to use CLT if $X_1, \dots, X_n \overset{iid}{}$ and n is "large"?

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

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

$X_1, \dots, X_{100} \overset{iid}{} \begin{cases} 1 & wp\ \frac{1}{2} \\ -1 & wp\ \frac{1}{2} \end{cases} \Rightarrow \mu = 0, \sigma^2 = 1 \Rightarrow \sigma = 1$

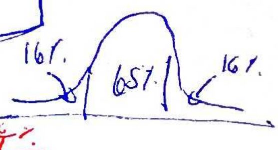
What is the probability of being more than or equal to 10 steps away from the origin after 100 steps?

$T = X_1 + X_2 + \dots + X_n \approx N(n\mu, (\sqrt{n}\,\sigma)^2) = N(0, 10^2)$

$P(T \geq 10 \text{ or } T \leq -10) = P(T \geq 10) + P(T \leq -10) =$

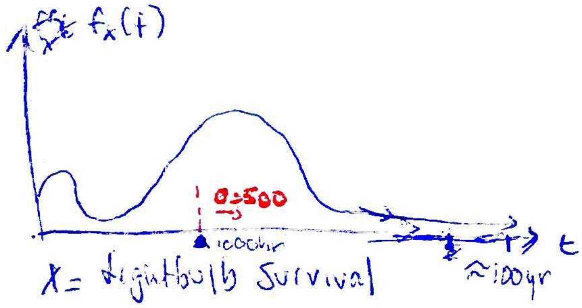
it does matter if it is $\geq$ or $\leq$

$= P(|T| \geq 10)$



16%

$= P\left(\frac{T-0}{10} \geq \frac{10-0}{10}\right) + P\left(\frac{T-0}{10} \leq \frac{-10-0}{10}\right) = P(Z \geq 1) + P(Z \leq -1) = 16\% + 16\% = 32\%$



If you get 50 bulbs, what is the prob the average lifetime is more than 1300 hrs?

$P(\bar{X} > 1300)$

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

by CLT

standardize

$\hookrightarrow P(\bar{X} > 1300) \approx P\left(\frac{\bar{X} - 100}{70.7} > \frac{1300-1000}{70.7}\right) = P(Z > 4.24)$

≈ 0

Z

Shipments are late 2% of the time.

In 10,000 orders, what is the probability more than 3% are late?

$P(\bar X > 0.03) \approx p$ $\qquad X_1, X_2, \dots, X_{1000} \overset{iid}{\sim} \text{Bern}(0.02)$

$\Rightarrow \mu = 0.02$

$\sigma = \sqrt{0.02(1-0.02)}$

$= 0.14$

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

$P(\bar x > 0.03) \approx P\left(\frac{\bar x - 0.02}{0.0014} > \frac{0.03 - 0.02}{0.0014}\right) = P(z > 7.14) \approx 0$

General

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

estimate

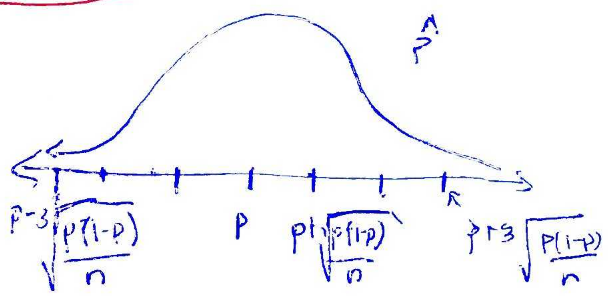
$\hat p = \bar x = \frac{1+1+0+0+0}{5} = 0.4$ sample of proportion denoted by $\hat p$

standard deviation

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

$\hat p = \bar X = \frac{1+1+0+0+0}{5} = 0.4 \approx p$ → is one realization of $\hat p$

special case of $\bar X$, if $\bar X \sim$ Bern, then



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

prob ↑

stat ↓

DONE WITH PROB, NOW STATS.

Statistical Inference

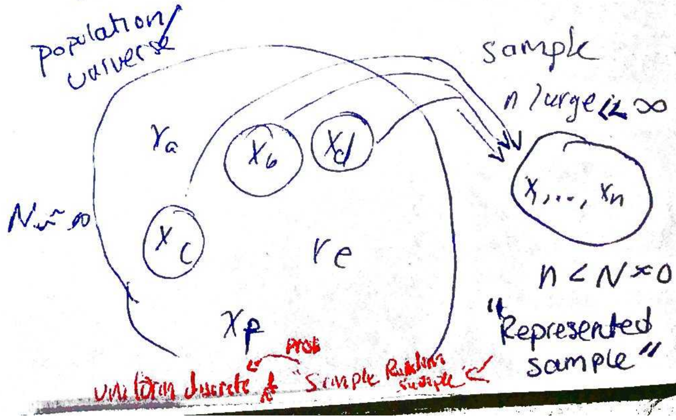
3 goals:

I) Estimate best guess of p ($\hat p$)

II) Estimate range/window ~~random~~ of p (confidence interval)

III) Test theories about p (hypothesis testing)

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



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

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

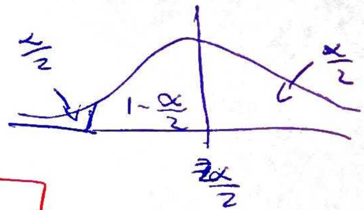
$$= P\left(-1 \le \frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \le 1\right) = P(-1 \le -Z \le 1) = P(1 \ge Z \ge -1).$$

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

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right]$$

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

$$P\left(p \in \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right]\right) = P\left(Z \in \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]\right)$$



$$= F\left(z_{\frac{\alpha}{2}}\right) - F\left(-z_{\frac{\alpha}{2}}\right) = \left(1-\frac{\alpha}{2}\right) - \left(\frac{\alpha}{2}\right) = \boxed{1-\alpha}$$

$$97.5\% = \int_{-\infty}^{z_{\frac{\alpha}{2}} = 2} f_Z(z)\,dz$$

unknown

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right]$$ WRONG, because p is unknown

when repeated ... gives you $1-\alpha$ "coverage" of p.

100 yrs of debate

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

2 sided

1-proportion confidence interval