Lecture 21

11/29/17

How to use CLT

If $X_1, \ldots, X_n \overset{iid}{\sim}$ and n is large

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

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

You take 100 random steps

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

What is the probability you are more than 10 steps away from where you start?

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

$P(|T| > 10) = P(T > 10 \text{ or } T < -10) = P(T > 10) + P(T < -10)$

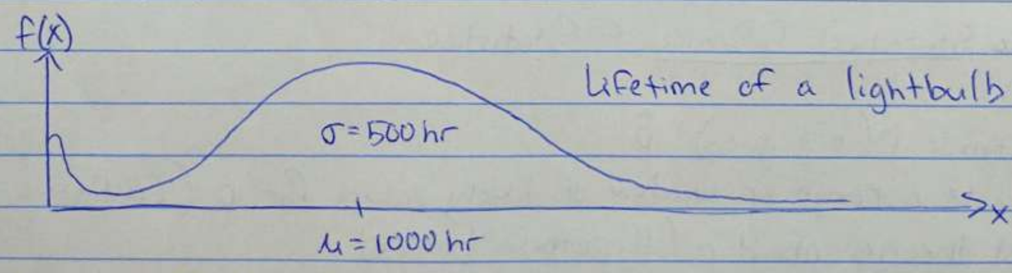
by CLT, $T \approx N(n\mu, (\sigma\sqrt{n})^2) = N(100 \cdot 0, (1 \cdot \sqrt{100})^2) = N(0, 10^2)$

$P(T > 10) + P(T < -10)$

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

$= P(Z > 1) + P(Z < -1)$

$= .16 + .16 = .32$



You get 50 lightbulbs

What is the probability the average lifetime is more than 1300 hours?

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

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

$P(\bar{X} > 1300) = P(\frac{\bar{X}-1000}{70.7} > \frac{1300-1000}{70.7}) \approx P(Z > 4.29) \approx 0$

Shipments are late 2% of the time. What is the probability in 10,000 shipments, more than 3% are late?

$X_1, \ldots, X_{10{,}000} \overset{iid}{\sim} \text{Bern}(0.02)$

$P(\bar{X} > 0.03)$

$\bar{X} \approx N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$, $\mu = 0.02$, $\sigma = \sqrt{0.02(1-0.02)} = 0.14$

$= N(0.02, (\frac{0.14}{\sqrt{10000}})^2) = N(0.02, 0.0014^2)$

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

$\bar{X}$ is the r.v. of the average

$\bar{x}$ is the realization

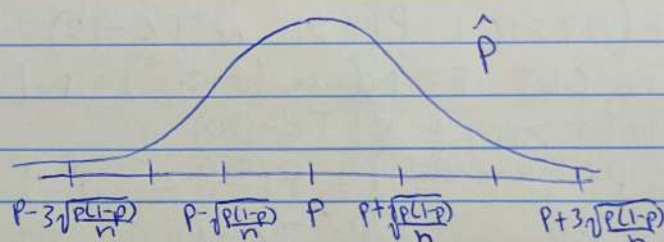
$\bar{x} = \frac{1+1+0+0+0}{5} = 0.4$ "Proportion"

$\bar{x} \in [0,1]$

$\hat{p}$ - Sample proportion

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

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



Statistical Inference $\subset$ Statistics

Goal:

① Estimate p (best guess): $\hat{p}$

② Create a range or window of likely values for p (Confidence interval)

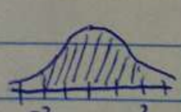
③ Test theories about p (Hypothesis theory)

$\hat{p} = \frac{\sum X_i}{n}$ - Simple random sample, otherwise bias

What is the probability: $P(p \in [\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}]) = P(\hat{p} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + \sqrt{\frac{p(1-p)}{n}})$

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

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



Define:

$Z_{\frac{\alpha}{2}} = F_Z^{-1}(1-\frac{\alpha}{2})$

$\alpha = 5\% \Rightarrow \frac{\alpha}{2} = 2.5\%$

$\Rightarrow 1 - \frac{\alpha}{2} = 97.5\%$

$\Rightarrow Z_{2.5\%} = 2$

Confidence Interval (CI)

$[\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}] \approx \hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$CI_{p,1-\alpha}$

Two sided, one proportion CI