WG: ~~Wednesday~~ MONDAY

CONFIDENCE INTERVALS. (INTERVAL estimation).

→ $X_1, \dots, X_n \sim f_\theta(x)$ FOR SOME parameter θ

MLE LEADS to $\hat{\theta}$ - NO INFORMATION ON INTRINSIC RELIABILITY.

WE want $[L(\text{Data}), U(\text{Data})]$ WHICH contains true θ WITH probability $(1-\alpha)$.

$\mathbb{P}\left(\theta \in [L(\text{Data}), U(\text{Data})]\right) = 1-\alpha$. (*) def CI.

EX: $X \sim \text{Bin}(n, p)$

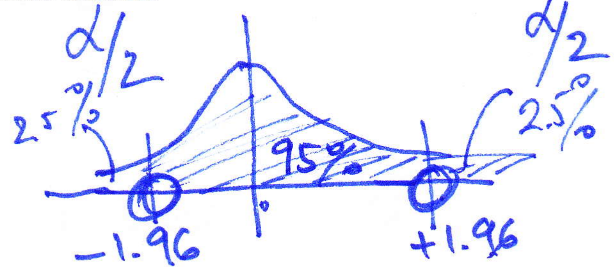
$\hat{p} = \frac{X}{n}$ - MLE ESTIMATOR OF Binomial parameter p.

$$\frac{\hat{p} - \mathbb{E}\hat{p}}{\sqrt{\text{Var}(\hat{p})}} \to \mathcal{Z}(0,1)$$

$\mathbb{E}\hat{p} = \mathbb{E}\left(\frac{X}{n}\right) = p.$

$\text{Var}\,\hat{p} = \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\text{Var}(X) = \frac{1}{n^2} \cdot n \cdot p(1-p) = \frac{p(1-p)}{n}$

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow Z(0,1)$$



$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) = 1-\alpha.$$

isolate p: too difficult quadratic inequality

It is justifiable to replace p by $\hat{p}$ in the denominator.

(as $\hat{p} \to p$ in some strong sense)

Hence

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\alpha/2}\right) = 1-\alpha$$

isolate p.

$$\mathbb{P}\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha$$

$$\Rightarrow CI_p = \left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

jbradic

HW6[redacted]

[21/Feb/2013:12:30:42 -0800] - p5880c-238474 - math64.ucsd.edu

MARGIN OF ERROR $\longmapsto$

is $\underbrace{1.96}_{2.} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (radius of CI)

How to choose sample size to get a $(1-\alpha)$-CI to have radius ε (i.e. to be reasonably sure that $\hat{p}$ is within ε of p).

$$\varepsilon = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\Rightarrow \quad n = \frac{4\hat{p}(1-\hat{p})}{\varepsilon^2}$$

CI for p gets shorter when p is near 0 or 1 (CLT is less reliable in such cases).

$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

unknown unknown.

$CI_\alpha(\mu) = ?$

$\hat{\mu} = \bar{X}$

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \to Z(0,1)$$

BUT WE DON'T KNOW σ.

Replace σ with $\hat{\sigma}$ IN THE DENOMINATOR.

random variable

$$\frac{\bar{X} - \mu}{\hat{\sigma} / \sqrt{n}} \sim t_{n-1}$$

Student t distribution

random variable

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$\mathbb{E}\,\hat{\sigma}^2 = \sigma^2$

$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})\right) = \frac{n-1}{n} \sigma^2$

$$\mathbb{P}\left(-t_{\alpha/2, n-1} \le \frac{\bar{X} - \mu}{\sqrt{\hat{\sigma}^2 / n}} \le t_{\alpha/2, n-1}\right) = 1 - \alpha$$

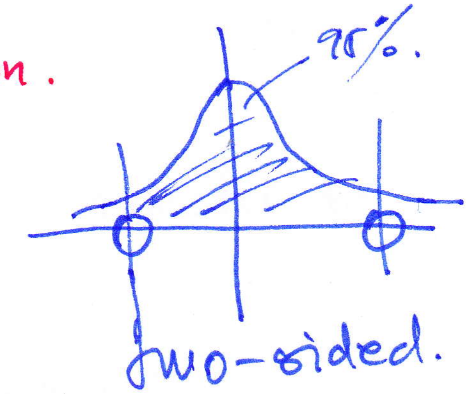
$\downarrow$ isolate μ

$$\mathbb{P}_\mu\left(\bar{X} - t_{\alpha/2, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right) = 1 - \alpha$$

$$CI = \left[\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} \cdot t_{\alpha/2, n-1},\ \bar{X} + t_{\alpha/2, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right)$$

$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

μ unknown, σ^2 unknown.



$CI_{\sigma^2} = ?$

$$\boxed{\frac{(n-1)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}}$$

$$\mathbb{P}\left(\chi^2_{\alpha/2,\, n-1} \le \frac{(n-1)\,\hat{\sigma}^2}{\sigma^2} \le \chi^2_{1-\alpha/2,\, n-1} \right) = 1-\alpha$$

$\downarrow$ isolate σ^2

$$\mathbb{P}\left(\sqrt{\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,\, n-1}}} \le \sigma^2 \le \sqrt{\frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2,\, n-1}}} \right) = 1-\alpha$$

$$CI = \left[\sqrt{\frac{(n-1)\,\sigma^2}{\chi^2_{1-\frac{\alpha}{2},\, n-1}}} \,,\, \sqrt{\frac{(n-1)\,\sigma^2}{\chi^2_{\alpha/2,\, n-1}}} \right]$$