Stat 231

October 19, 2016.

No Tutorial today

Next week → 6-00 clock
Tutorial

Banerjee → STP 105.

Cyntha ← 3-30 (DC 1351?)

Interval Estimation

Method of Sampling Distributions.

- Definitions
- Gaussian with known variance.
- Binomial estimation using the CLT.
- How to choose the "right" sample size?

θ = unknown parameter (attribute of interest).

examples : μ, σ^2, $\tilde{x}$, Max, Min, . . .

θ $\{y_1, \ldots y_n\} \longrightarrow$ SAMPLE

$\hat{\theta}$ = <u>Estimate</u> (MLE)
= $g(y_1, \ldots y_n)$.

Examples
$\bar{y}$, s^2, $\hat{\sigma}^2$
$\hat{x}$, . . .

$\hat{\theta}$ would be different from sample to sample

$\hat{\theta}$ is thought of as an outcome of a random variable.

$\tilde{\theta}$ = r.v. = ESTIMATOR OF θ.

An estimator is not a #, but a r.v. (Notation: $\tilde{\theta}$, $\bar{Y}$, S^2)

The distribution of $\tilde{\theta}$ = SAMPLING DISTRIBUTION OF THE ESTIMATOR.

We use the sampling distribution (or a function of it) to construct our interval estimates.

~~Definit~~

Definition: A $100p\%$ Confidence interval for θ is the estimate $[l, u]$ of the random ~~variable~~ interval $[L, U]$, such that

$$P(L < \theta < U) = p.$$

$p = 0.95$

If the experiment is repeated many times, 95% of the intervals constructed would contain θ.

<u>Example</u> 1

GAUSSIAN PROBLEM WITH KNOWN VARIANCE.

We are interested in finding a 99% C.I for starting salaries of UW Math graduates

We know the population s.d. of starting salaries = 10,000. = σ.

A sample of 25 UW grads are taken $\{y_1, \ldots y_n\}$

$\bar{y} = 75{,}000$

Step 1 Set up the model.

$$Y_i \sim G(\mu, 10{,}000)$$

$i = 1, \ldots, 25$

Y_i's independent.

Step 2: Find the MLE of μ.

$$\hat{\mu} = \bar{y} = 75{,}000$$

Step 3: Identify the sampling distribution of $\bar{Y}$

From ~~\$~~ STAT 230,

SAMPLING DIST$^{\underline{N}}$

$$\boxed{\bar{Y} \sim G\left(\mu, \frac{10{,}000}{\sqrt{25}}\right)}$$

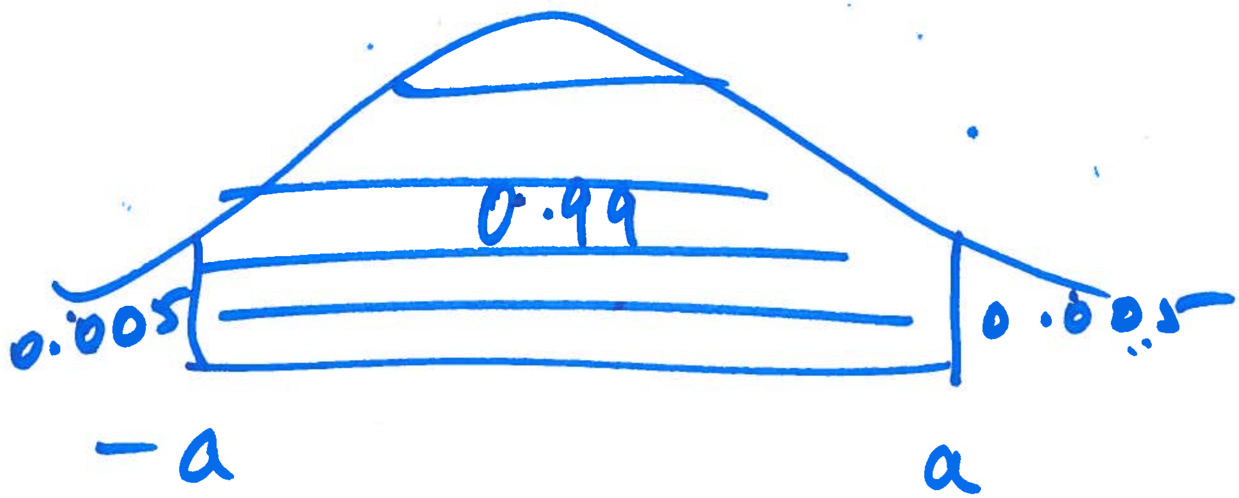
$$\left\{ X_1, \ldots X_n \sim N(\mu, \sigma^2) \quad \bar{X} \sim N(\mu, \sigma^2/n) \right\}$$

Step 4 Construct the pivotal distribution

$$\boxed{\frac{\bar{Y} - \mu}{2000} = Z \sim N(0,1)}$$

PIVOTAL DISTRIBUTION

Step 5 Find the end points of your pivotal distribution.



$a = 2.58$ (Check!)

Step 6: Combine Step 4 and 5 to find the coverage interval.

$$P(-2.58 \le Z \le 2.58) = 0.99$$

(Step 5)

$$P\left(-2.58 \le \frac{\bar{Y} - \mu}{2000} \le 2.58\right) = 0.99$$

(Step 4)

$$\mu \ge \bar{Y} - 2.58 \times 2000$$

$$\mu \le \bar{Y} + 2.58 \times 2000$$

Coverage Interval.

$$\left[\bar{Y} - 2.58 \times 2000, \quad \bar{Y} + 2.58 \times 2000 \right]$$

Step 7 : Estimate the coverage interval. using your sample.

C. I

$$\left[\bar{y} - 2.58 \times 2000, \quad \bar{y} + 2.58 \times 2000 \right]$$

C. I $\boxed{75,000 \pm 2000 \times 2.58}$

$10,000/\sqrt{25}$

General formula

Gaussian data with known variance.

C.I: $\boxed{\bar{y} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}}$

Can we choose the MOE for this problem to be of any specified length?

For the previous problem, Can we make sure that the C.I is

$$\boxed{\bar{y} \pm 1000}$$

?

$$\bar{y} \pm \frac{z^{\alpha} \cdot \sigma}{\sqrt{n}}$$

$$\frac{z^{\alpha}\sigma}{\sqrt{n}} = 1000$$

$$n = \left(\frac{z^{\alpha} \cdot \sigma}{1000}\right)^2$$

Question Gaussian with known variance.

$n = 1000.$

~~$\bar{y} \pm a.$~~

$\bar{y} \pm a.$

New margin of error $= a/2$

What sample size should you choose?

(a) $n = 250$

(b) $n = 500$

(c) $n = 2000$

(d) $\boxed{n = 4000}$ $\rightarrow$ 73%

Case II Binomial problem with n large.

$$Y \sim \text{Bin}(n, \theta)$$

θ = probability of success.

Sample $n = 200$ $y = 30$.

$$\hat{\theta} = 30/200 = 0.15$$

$\parallel$

Sample proportion

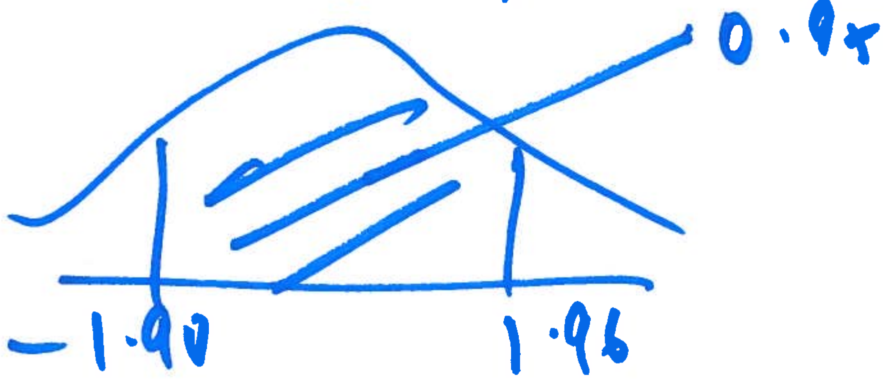
Theorem: For the Binomial problem

$$\frac{\hat{\theta}-\theta}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}} = Z \sim N(0,1)$$

(by using the CLT)

Find the 95% C.I for θ.

$$P(-1.96 \le Z \le 1.96) = 0.95$$



$$P\left(-1.96 \le \underbrace{\frac{\tilde{\theta} - \theta}{\sqrt{\frac{\hat{\theta}(1-\tilde{\theta})}{n}}}} \le 1.96\right) = 0.95$$

$$\theta \le \tilde{\theta} + 1.96\sqrt{\frac{\tilde{\theta}(1-\hat{\theta})}{n}}$$

$$\theta \ge \tilde{\theta} - 1.96\sqrt{\frac{\tilde{\theta}(1-\hat{\theta})}{n}}.$$

Coverage Interval

$$\boxed{\tilde{\theta} \pm 1.96\sqrt{\frac{\tilde{\theta}(1-\hat{\theta})}{n}}}$$

Confidence Interval

$$\boxed{\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$$

$$0.15 \pm 1.96\sqrt{\frac{0.15 \times 0.85}{200}}$$

C.I : $$\boxed{\hat{\theta} \pm z^{\alpha}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$$

where $\hat{\theta}$ = Sample proportion