STAT 231

October 17, 2016

Roadmap

INTERVAL ESTIMATION

* Likelihood based intervals

* Confidence Interval

    * Gaussian with Known variance

    * Binomial Distribution.

Objective

θ = unknown population parameter.

Model: $Y_i \sim f(y_i; \theta)$

$i = 1, \dots n$

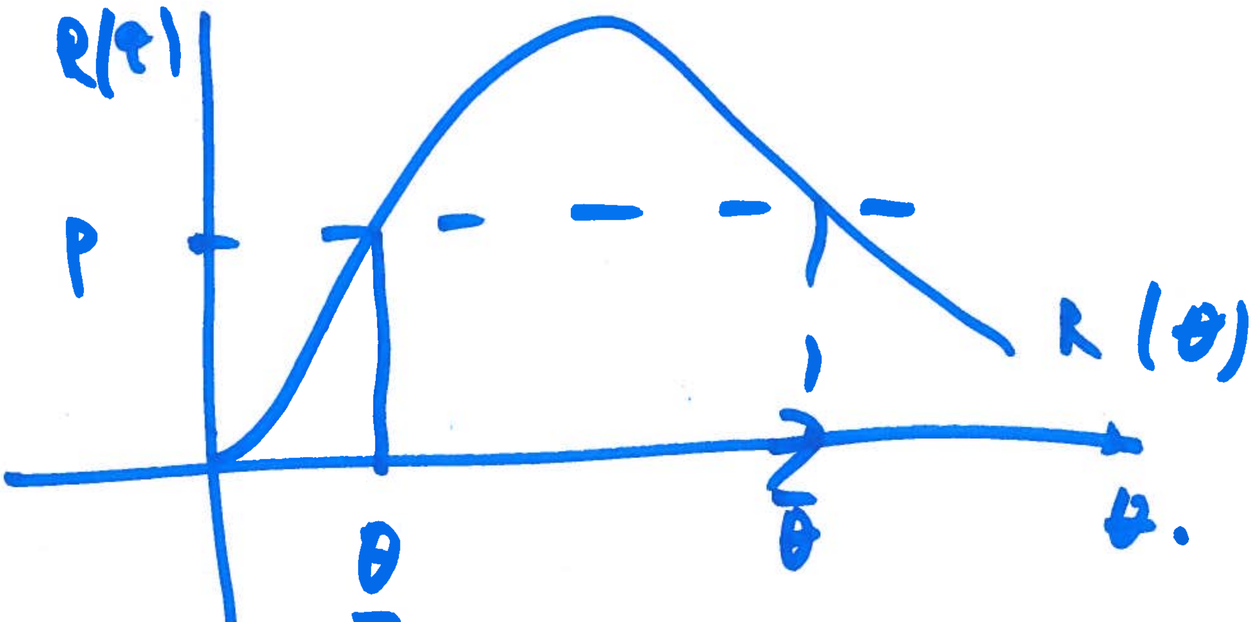
Data: $\{y_1, \dots y_n\}$

OBJECTIVE: Based on your sample, we want to estimate the interval that contain θ with a high probability. $[l, u]$

METHOD I : LIKELIHOOD FUNCTION.

$100p\%$ l.i. for θ

$= \{ \theta : R(\theta) \geq p \}$



$[\underline{\theta}, \bar{\theta}] = 100p\%$ likelihood interval.

CONVENTIONS

If θ Θ lies in the 50% → very plausible

10% → plausible.

outside 10% — implausible

outside 1% → very "

Example: Suppose θ = unknown parameter = prob. of success in a Binomial distribution

$$Y \sim Bin(n, \theta).$$

$n = 500$, $y = \#$ of successes
$= 200.$

Questions

(i) Find the 15% likelihood interval for θ.

(ii) Is $\theta = 0.5$ plausible?

(i) $\{\theta : R(\theta) \geq 0.15\} = ?$

What is $R(\theta) = ?$

$$R(\theta) = L(\theta) / L(\hat{\theta})$$

$\hat{\theta} = \text{MLE}$

$$\boxed{L(\theta) = {}^{500}C_{200}\, \theta^{200} (1-\theta)^{300}}$$

$$\boxed{\hat{\theta} = 200/500 = 0.4} \longrightarrow \text{from MDT 1}$$

$$R(\theta) = \frac{\cancel{{}^{500}C_{200}}\, \theta^{200}(1-\theta)^{300}}{\cancel{{}^{500}C_{200}}\,(0.4)^{200}(0.6)^{300}}$$

To find the 15% l. i

$R(\theta) \geq 0.15$

15% Likelihood Interval.

$$\boxed{\frac{\theta^{200}(1-\theta)^{300}}{(0.4)^{200}(0.6)^{300}} \geq 0.15}$$

(*)

To find whether $\theta = 0.5$ is plausible, plug $\theta = 0.5$ in $*$. and see whether it is $\geqslant 0.1$.

METHOD II : METHOD OF SAMPLING DISTRIBUTIONS:

ESTIMATES

Insight: all numerical measures ~~are~~ can be thought of as outcomes of a random variable.

ESTIMATOR

Example:

Suppose a sample of 36 observation are drawn from a Gaussian dist-ribution with mean μ

$\bar{y} = 80$ variance = 49

Use this data set to construct a 95% C.I. for μ

Confidence Interval.

Step 1: Find the MLE for μ.

$\bar{y} = 80 = $ MLE for μ.

$Y_1, \ldots Y_n \sim G(\mu, \sigma)$

$\bar{Y} \sim G(\mu, \sigma/\sqrt{n})$

(STAT 230)

$\bar{Y}$ = estimator

$\bar{y}$ = estimate

$$\bar{Y} \sim G\left(\mu, \frac{7}{\sqrt{36}}\right)$$

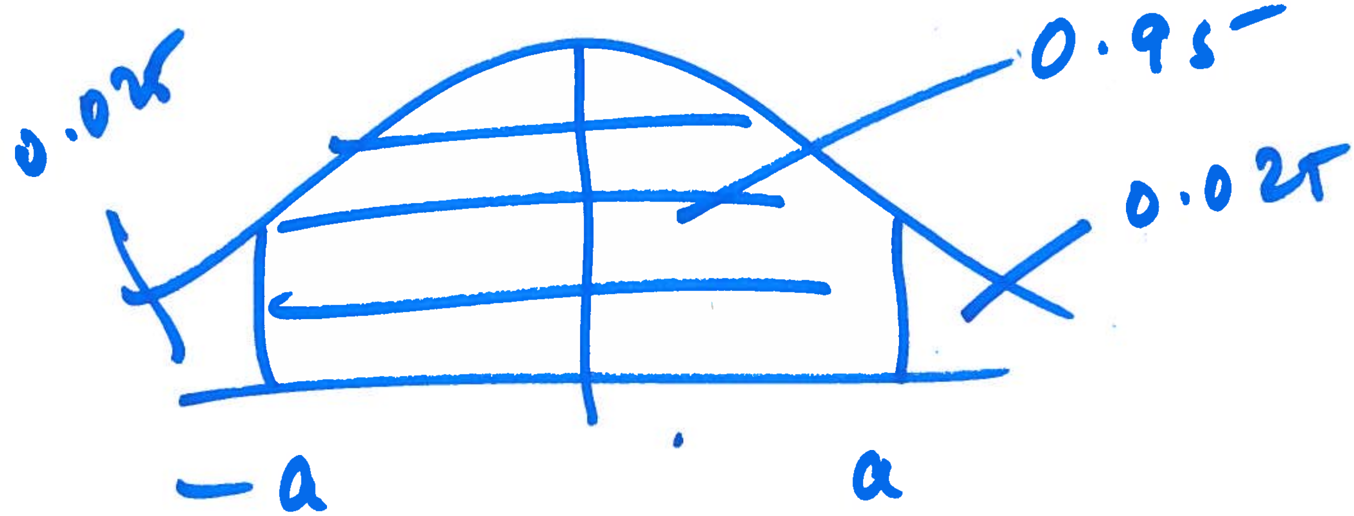
$$\bar{Y} \sim G\left(\mu, \frac{7}{6}\right)$$

$$\boxed{\frac{\bar{Y} - \mu}{7/6} = Z \sim G(0,1)}$$

This follows from.

$$Y \sim G(\mu, \sigma) \qquad \frac{Y - \mu}{\sigma} = Z$$

Find the 95% interval for Z.



$\boxed{a = 1.96}$ ⟶ Check!!

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

$$= P\left(-1.96 \leq \frac{\bar{X}-\mu}{7/6} \leq 1.96\right) = 0.95$$

(Right hand Inequality)

$$\mu \geq \bar{Y} - 1.96 \cdot 7/6$$

$$\mu \leq \bar{Y} + 1.96 \cdot 7/6$$

$$\left[\bar{Y} - 1.96\ 7/6 \ldots\ \bar{Y} + 1.96 \cdot 7/6\right]$$

Coverage Interval: The random interval that contains μ with a high probability (95%)

Your estimate for this interval.

$$\left[\bar{y} - 1.96 \cdot 7/6 \,,\; \bar{y} + 1.96\, 7/6\right]$$

↑

CONFIDENCE INTERVAL

$$\left[80 - 1.96\, 7/6 \,,\; 80 + 1.96\, 7/6\right]$$

Given my sample, a 95% C.I is

$$\left[80 \pm 1.96\, 7/6\right]$$

The confidence is <u>AN ESTIMATE</u> of the actual <u>random interval</u> (COVERAGE INTERVAL) that contains θ with a high probability.

The example is unrealistic because σ^2 typically is unknown.

Example

500 US voters are interviewed at random.

220 of them plan to vote for Trump.

Find a 95% C.I for θ = proportion of Trump voters.

$\hat{\theta} = 220/500 = 0.44$

$\tilde{\theta}$: What is the distⁿ of the eshmator?

By the CLT (Central Limit Theorem), if n is large.

$$\boxed{\frac{\tilde{\theta} - \theta}{\sqrt{\tilde{\theta}(1-\tilde{\theta})/n}} = Z \sim G(0,1)}$$

The 95% C.I for the Binomial problem:

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

$$0.44 \pm \underbrace{1.96\sqrt{\frac{0.44 \times 0.56}{500}}}_{\text{MOE}}$$