Interval Estimation 1

Tues. Jan 14

Point estimation: Estimate θ with $\hat{\Theta}_n(X_1 \ldots X_n)$
A single number

Interval Estimation : Guess that θ is in some interval

between L (lower confidence limit) & U (upper confidence limit)
(Margin of error)

want $P(L < \theta < U)$ high, like 0.95, 0.99

Ex $X_1 \ldots X_n \overset{iid}{\sim}$ Bernoulli (θ)

$P(X_i = 1) = E(X_i)$ = prob alive 5 years after diagnosis

Say $\frac{140}{200}$ lived 5+ years

$\hat{\theta} = 0.7$ point estimate

CLT says $Z_n = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0,1)$

$\mu = \theta$, $\sigma^2 = \theta(1-\theta)$ θ is unknown

Modified CLT says

$$Z_n = \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\hat{\sigma}_n} \xrightarrow{d} Z \sim N(0,1),$$

where $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$

Here, take $\hat{\sigma}_n^2 = \bar{X}_n(1-\bar{X}_n)$ consistent by LLN & continuous mapping

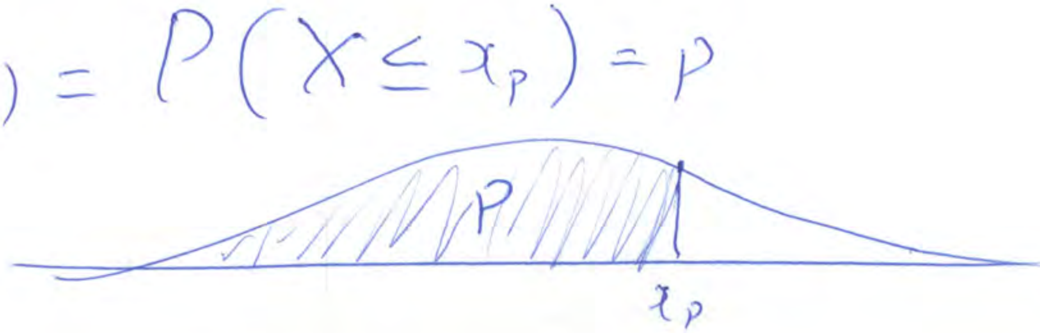
---

Quantile

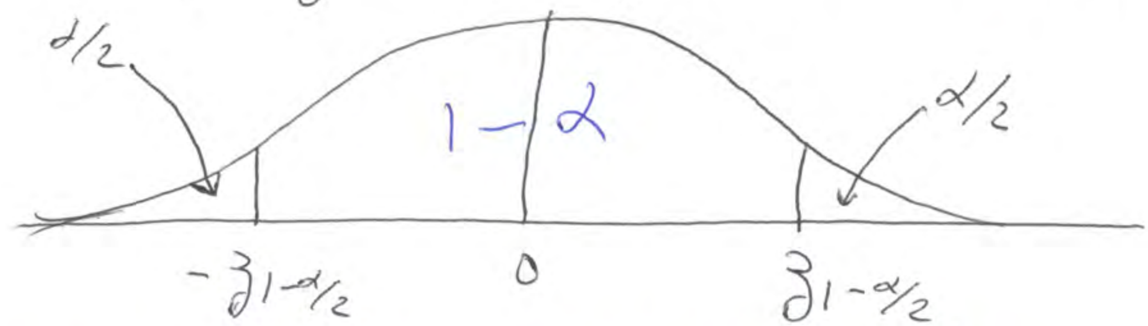
The p quantile or pth quantile of a distribution is the point with p of the probability at or below it.

If X is a RV with cdf $F_X(x)$, the pth quantile of dist. of X is the point x_p satisfying

$$F_X(x_p) = P(X \le x_p) = p$$



Let α is small number, like $\alpha = 0.05$ or $\alpha = 0.01$. So Denoting by $z_{1-\alpha/2}$ the $1-\frac{\alpha}{2}$ quantile of the standard normal



$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha$$

like $1-\alpha = 0.95$ or 0.99

We will a $(1-\alpha)*100\%$ confidence interval for μ, like a 95% confidence interval, or 99% confidence interval

$$1-\alpha = P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\alpha/2}\right)$$

$$\approx P\left(-z_{1-\frac{\alpha}{2}} < Z_n < z_{1-\alpha/2}\right)$$

$$= P\left(-z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} < z_{1-\frac{\alpha}{2}}\right)$$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\hat{\sigma}_n} < z_{1-\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}} < \bar{X}_n - \mu < z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}\right)$$

$$= P\left(-\bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}} < -\mu < -\bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}} > \mu > \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}\right)$$

$$= P\Big(\underbrace{\bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}}_{L} < \mu < \underbrace{\bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}}_{U}\Big)$$

or $\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}$

For Bernoulli survival, want 95% confidence interval

$$\bar{x}_n \pm \frac{\hat{\sigma}}{\sqrt{n}} z_{1-\frac{\alpha}{2}} = \bar{x}_n \pm \sqrt{\frac{\bar{x}(1-\bar{x})}{200}} * 1.96$$

$$= 0.7 \pm \sqrt{\frac{.7 * .3}{200}} * 1.96$$

$$= 0.7 \pm 0.0635$$ 95% margin of error

$$= (0.636, 0.764)$$ We are "95% confident" that the 5-year survival rate is between .64 & .76

Does this mean $P(0.636 < \theta < 0.764) = 0.95$? **NO** (5)

Constants

"The estimated survival probability is 70%. This is expected to be accurate within 6 percentage points, 19 times out of 20."

(Comes from 0.7 ± 0.06)

In a random sample of 49 Toronto homes more than 50 years old, the sample mean ~~of 49 houses~~ lead concentration in tap water after running the water for one minute was 0.657 micrograms per litre, with a standard deviation 0.21 mcg/l. Give a 95% CI for average amount of lead in TO tap water.

CI is $\bar{x}_n \pm z_{1-\alpha/2} \frac{s_n}{\sqrt{n}} = 0.657 \pm 1.96 \frac{0.21}{\sqrt{49}}$

$= 0.657 \pm 1.96 \cdot 0.03 = \cancel{(0.654, 0.660)}$

$(0.598, 0.716)$

Basketball

Random sample of young adults warms up then shoots from the free throw line (< 5 m) until making a basket X_i = # misses

Assumption: Conditionally on shooter's ability making or missing is like tossing a coin (independent).

Model: $X_1 \ldots X_n \stackrel{iid}{\sim}$ Geometric(θ)

Disagree

2-stage model

1. Choose accuracy θ from a population with expected value μ_θ.
2. Shoot till you hit, record X_i = # of misses

~~$P(X_i = 0) = (1-\theta)^0 \theta = \theta$~~

$P(X_i = 0 \mid \Theta_i = \theta) = \theta$ so

Law of total probability

$$P(X_i = 0) = \int_0^1 P(X_i = 0 \mid \Theta_i = \theta_i) f_\Theta(\theta_i)\, d\theta_i$$

$$= \int_0^1 \theta f_\Theta(\theta_i)\, d\theta_i = E(\Theta) = \mu_\theta$$