8. Confidence Intervals

Confidence Interval for the Kiss Example

And the question is, why?

[INAUDIBLE]

Exactly.

Because it depends on p, or p factorial,

just an exclamation mark.

So it depends on p.

Confidence interval?

- For a fixed $\alpha \in (0,1)$, if $q_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$, then with probability $\simeq 1-\alpha$ (if n is large enough !),

$$\bar{R}_n \in \left[p - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, p + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

- It yields

$$\lim_{n\to\infty} \mathbb{P}\left(\left[\bar{R}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right] \ni p\right) = 1-\alpha$$

- But this is **not** a confidence interval because it depends on p!

- To fix this, there are 3 solutions.

21/61

And I cannot do that, because if you give me data,

I can not build error bars, because there's this p here.

So what do I do?

Well, good news-- there's three ways to fix this.

18:45 / 18:52 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Random or Deterministic?

4/4 points (graded)

As in the video above, let $R_1, \ldots, R_n \stackrel{iid}{\sim} \text{Ber}(p)$ for some unknown parameter p. We estimate p using the estimator $\hat{p} = \overline{R}_n = \frac{1}{n}\sum_{i=1}^{n} R_i$.

For a fixed number α, after applying the CLT (and doing some algebra), we obtained

$$\lim_{n\to\infty} \mathbf{P}\left(\left[\overline{R}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \overline{R}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right] \ni p\right) = 1 - \alpha.$$

Which of the quantities in the equation above is random and which is deterministic? (Choose one for each column.)

$\overline{R}_n$:	n :	$q_{\alpha/2}$:	p :
(●) random	( ) random	( ) random	( ) random
( ) deterministic	(●) deterministic	(●) deterministic	(●) deterministic
✔	✔	✔	✔

(The submit button is activated only after you have answered each question.)

Solution:

- $\overline{R}_n = \frac{\sum_{i=1}^{n} R_i}{n}$ is function of the random variables R_i, and hence is random.
- n is the sample size, a deterministic number.
- $q_{\alpha/2}$ is a number given a fixed α, hence deterministic.
- p is the unknown parameter, a number, hence deterministic.

Remark 1: Once we substitute a realization for $\overline{R}_n$ (e.g. from data), the expression $\left[\overline{R}_n - \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \overline{R}_n + \frac{q_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right] \ni p$ becomes deterministic since all involved quantities are deterministic.

Remark 2: The unknown parameter p is deterministic in the classical (frequentist) approach. In the course *6.431x, Probability-the Science of Uncertainty and Data*, we have seen that in the Bayesian approach, p is modeled as a random variable. We will revisit Bayesian statistics from a different perspective later in this course.

Submit

You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Foundation of Inference:Lecture 4: Parametric Estimation and Confidence Intervals / 8. Confidence Intervals

Add a Post

Show all posts ▼ by recent activity ▼

Confusing?

I thought I understood confidence intervals fairly well, but found the math manipulations very confusing in this video. Hopefully, things will become ... 4

Learn About Verified Certificates