EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the Privacy Policy.





Lecture 4: Parametric Estimation

Course > Unit 2 Foundation of Inference > and Confidence Intervals

> 8. Confidence Intervals

Audit Access Expires Dec 24, 2019

You lose all access to this course, including your progress, on Dec 24, 2019.

Upgrade by Nov 4, 2019 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

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

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

- ▶ But this is **not** a confidence interval because it depends on p!
- ► To fix this, there are 3 solutions.

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.



Video

Download video file

Transcripts

<u>Download SubRip (.srt) file</u> Download Text (.txt) file

Random or Deterministic?

4/4 points (graded)

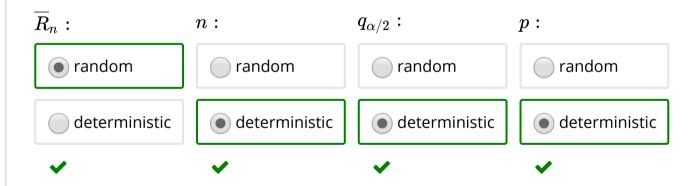
As in the video above, let $R_1,\ldots,R_n\stackrel{iid}{\sim}\mathsf{Ber}\,(p)$ for some unknown parameter p. We estimate p using the estimator

$$\hat{p} = \overline{R}_n = rac{1}{n} \sum_{i=1}^n R_i.$$

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

$$\lim_{n o\infty}\mathbf{P}\left(\left[\overline{R}_{n}-rac{q_{lpha/2}\sqrt{p\left(1-p
ight)}}{\sqrt{n}},\overline{R}_{n}+rac{q_{lpha/2}\sqrt{p\left(1-p
ight)}}{\sqrt{n}}
ight]
ightarrow p
ight)=1-lpha.$$

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



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

Solution:

- ullet $\overline{R}_n = rac{\sum_{i=1}^n R_i}{n}$ is function of the random variables R_i , and hence is random.
- ullet n is the sample size, a deterministic number.
- $q_{lpha/2}$ is a number given a fixed lpha, 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

$$\left[\overline{R}_n - \frac{q_{\alpha/2}\sqrt{p\left(1-p\right)}}{\sqrt{n}}, \overline{R}_n + \frac{q_{\alpha/2}\sqrt{p\left(1-p\right)}}{\sqrt{n}}\right] \ni p \text{ becomes deterministic since all involved quantities are }$$

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

1 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

