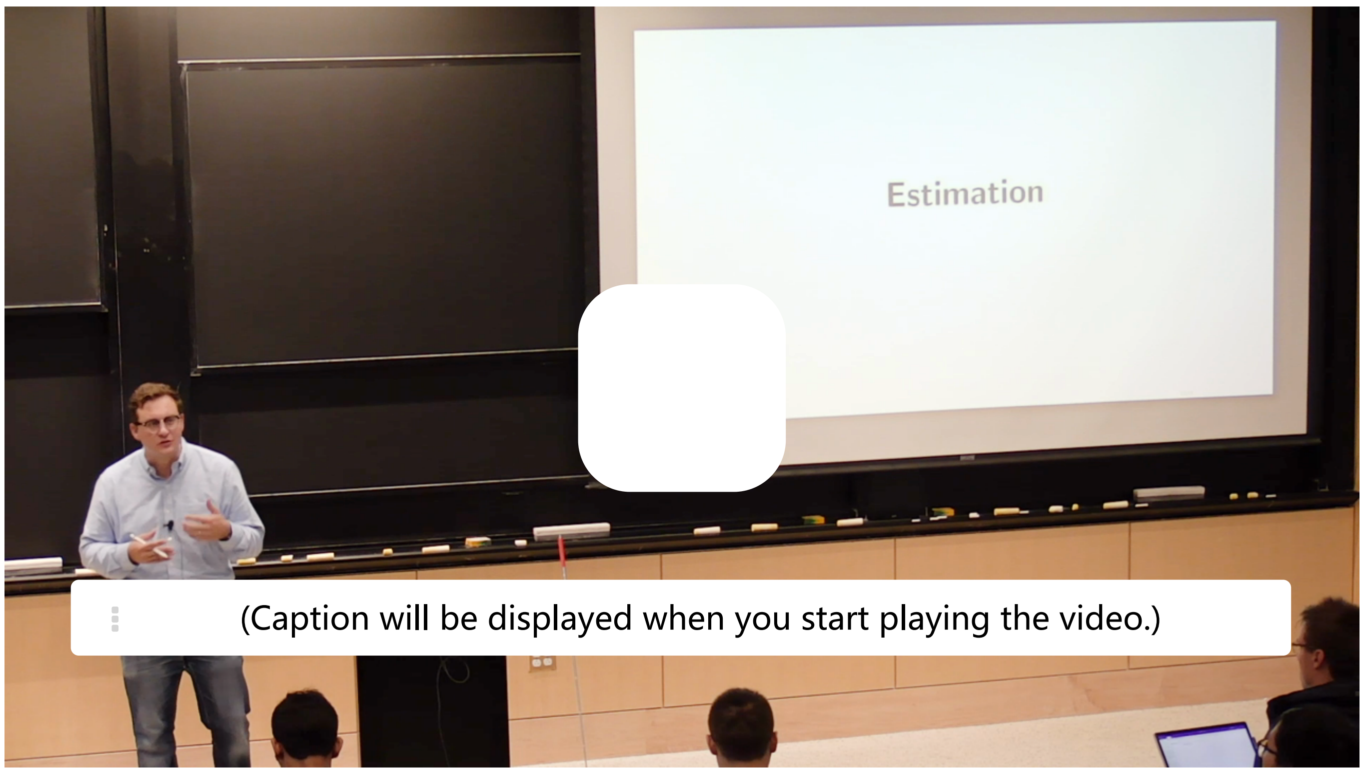
2. Statistics, Estimators, Consistency, and Asymptotic Normality

Statistics, Estimators, Consistency, and Asymptotic Normality



Start of transcript. Skip to the end.

OK.

Welcome back.

Last week, we laid the framework for our trinity

of statistical inference, which are namely

estimation, confidence interval, and hypothesis testing.

And the foundation that we laid was this idea

of statistical modeling, this idea

视频

下载视频文件

字幕

下载 SubRip (.srt) file

下载 Text (.txt) file

Which Statistics are Estimators?

1/1 point (graded)

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} P_\theta$ where the distribution P_θ depends on an unknown parameter $\theta \in \mathbb{R}$. Which of the following statistics are considered **estimators** ?

(Choose all that apply.)

- [ ] θ
- [x] 4.2 ✔
- [x] $\sum_{i=1}^{n} i^2 X_i^i$ ✔
- [x] $\frac{1}{n}\sum_{i=1}^{n} X_i$ ✔
- [ ] $\frac{1}{n}\sum_{i=1}^{n} X_i - \theta$

✔

Solution:

Recall that heuristically, a statistic is a function of the data that can be easily computable, and an estimator is a statistic whose expression does not depend on the unknown parameter θ.

Note that the first and last choices, θ and $\frac{1}{n}\sum_{i=1}^{n} X_i - \theta$ both have some explicit dependence of θ, so they cannot be estimators.

On the other hand, the remaining expressions 4.2, $\sum_{i=1}^{n} i^2 X_i^i$, and $\frac{1}{n}\sum_{i=1}^{n} X_i$ only depend on $X_1, \ldots, X_n$ (and not θ), so they are indeed estimators.

Remark: The second estimator, 4.2, is potentially a very poor choice as it does not depend on the data set. But according to the definition, it is still considered an estimator.

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

Consistency of an Estimator

1/1 point (graded)

An estimator $\hat{\theta}_n$ is **weakly consistent** if $\lim_{n\to\infty} \hat{\theta}_n = \theta$, where the convergence is in probability.

Suppose that in the previous problem the unknown parameter θ is the common mean of $X_1, \ldots, X_n$. Assume that $\theta \neq 4.2$. Which of the following is a weakly consistent estimator for θ? (Choose all that apply.)

- [ ] θ
- [ ] 4.2
- [ ] $\sum_{i=1}^{n} i^2 X_i^i$
- [x] $\frac{1}{n}\sum_{i=1}^{n} X_i$ ✔
- [ ] $\frac{1}{n}\sum_{i=1}^{n} X_i - \theta$

✔

Solution:

By the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1] = \theta$ in probability, so this is the correct choice.

From the previous question, the first and last choice, θ and $\frac{1}{n}\sum_{i=1}^{n} X_i - \theta$, are not even estimators, so these options are incorrect. Since $\theta \neq 4.2$, this estimator cannot be consistent. Finally, there is no guarantee that $\sum_{i=1}^{n} i^2 X_i^i$ converges to θ. In fact, for many choices of distribution, this statistic will diverge to ∞.

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

Quantifying Consistency

0/1 point (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber(p)}$. Let $\overline{X}_n$ be the estimator given by $\frac{1}{n}\sum_{i=1}^{n} X_i$.

What is the smallest constant c such that

$$n^c \left(\overline{X}_n - p\right) = n^c \left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right)$$

does **not** converge to 0 almost surely as $n \to \infty$?

1

✖ Answer: .5

Solution:

Let $\sigma = \sqrt{p(1-p)}$ denote the common standard deviation of $X_1, \ldots, X_n$. By the central limit theorem,

$$\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - p\right) = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right) \to N(0,1)$$

where the convergence is in distribution. As a result, we see that for $c < 1/2$,

$$n^c\left(\overline{X}_n - p\right) = \frac{\sigma}{n^{1/2-c}}\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - p\right) \approx \frac{\sigma}{n^{1/2-c}} N(0,1) \to 0$$

almost surely as $n \to \infty$. Hence, $c = 1/2$ is the smallest possible value of c such that

$$n^c\left(\overline{X}_n - p\right) = n^c\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right)$$

does **not** converge to 0 almost surely as $n \to \infty$.

Remark: As defined in the third video in this section, this implies that the estimator $\overline{X}_n$ is $\sqrt{n}$-consistent. This means that the estimator $\overline{X}_n$ converges to the true parameter at a relatively fast rate, so this gives us something stronger than just consistency.

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

讨论

显示讨论

主题： Unit 2 Foundation of Inference:Lecture 4: Parametric Estimation and Confidence Intervals / 2. Statistics, Estimators, Consistency, and Asymptotic Normality

认证证书是什么?