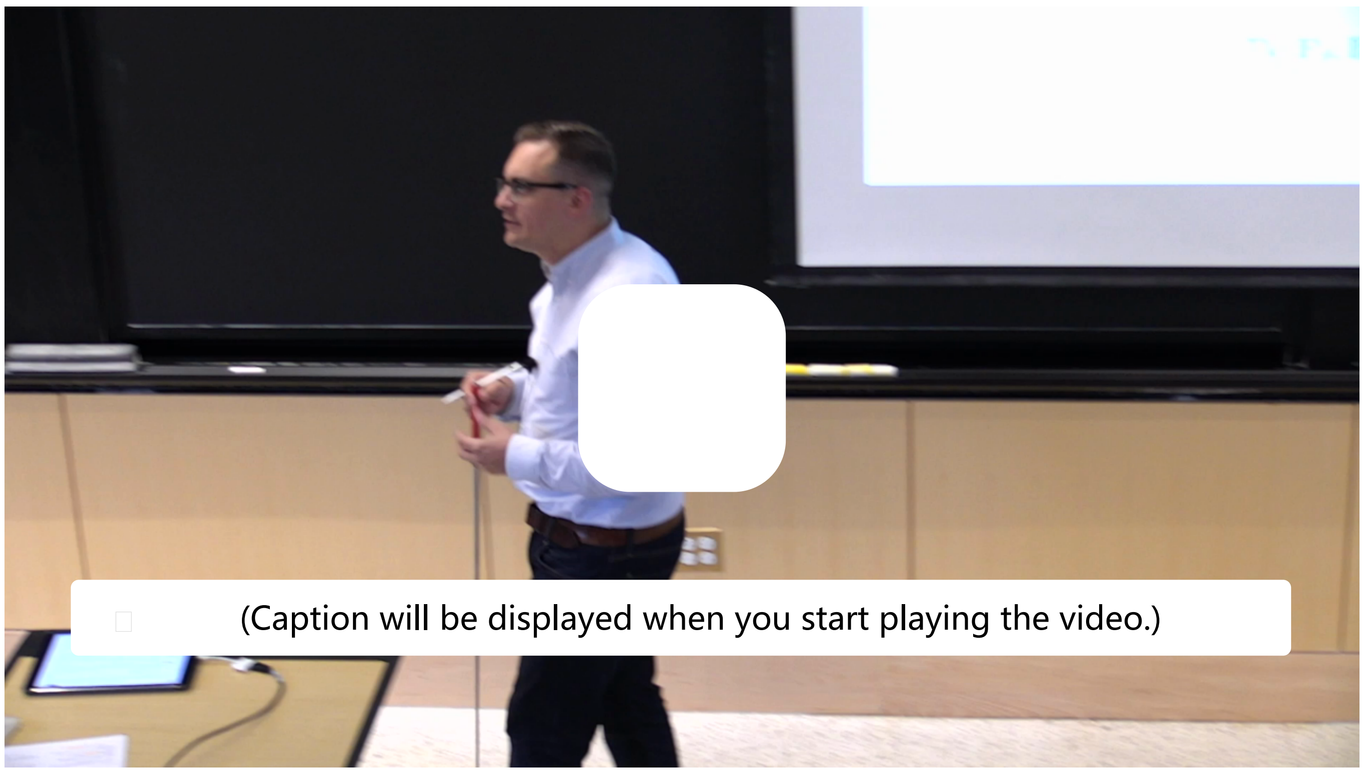
4. Introduction to Total Variation Distance

Definition of Total Variation Distance

Start of transcript. Skip to the end.



So, arguably, when you're trying to build a statistical method,

you need to understand what your goal is, right?

So if you're trying to understand

what it is to learn a distribution,

that's what we're trying to do.

We're trying to estimate the distribution,

but let's start from the basics and say, what is it we

视频

下载视频文件

字幕

下载 SubRip (.srt) file

下载 Text (.txt) file

Interpreting Total Variation Distance

1/1 point (graded)

Recall from lecture that the **total variation distance** between two probability measures $\mathbf{P}_\theta$ and $\mathbf{P}_{\theta'}$ with sample space E is defined by

$$\mathrm{TV}\left(\mathbf{P}_\theta, \mathbf{P}_{\theta'}\right) = \max_{A \subset E} \left|\mathbf{P}_\theta\left(A\right) - \mathbf{P}_{\theta'}\left(A\right)\right|$$

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbf{P}_{\theta^*}$ where $\theta^* \in \mathbb{R}$ is an unknown parameter. You construct a statistical model $\left(E, \{\mathbf{P}_\theta\}_{\theta \in \mathbb{R}}\right)$ for your data. By analyzing your data, you are able to produce an estimator $\hat{\theta}$ such that the distributions $\mathbf{P}_{\hat{\theta}}$ and $\mathbf{P}_{\theta^*}$ are close in **total variation distance**. More precisely, you know that

$$\mathrm{TV}\left(\mathbf{P}_{\hat{\theta}}, \mathbf{P}_{\theta^*}\right) \leq \epsilon,$$

where ϵ is a very small positive number.

Which of the following can you conclude about the distributions $\mathbf{P}_{\hat{\theta}}$ and $\mathbf{P}_{\theta^*}$? (Choose all that apply.)

- [x] Let A be an event. Then $\left|\mathbf{P}_{\theta^*}\left(A\right) - \mathbf{P}_{\hat{\theta}}\left(A\right)\right| \leq \epsilon$.
- [x] Let $X \sim \mathbf{P}_{\theta^*}$, let $Y \sim \mathbf{P}_{\hat{\theta}}$, and suppose $a, b \in \mathbb{R}$ where $a \leq b$. Then $\left|\mathbf{P}_{\theta^*}\left(a \leq X \leq b\right) - \mathbf{P}_{\hat{\theta}}\left(a \leq Y \leq b\right)\right| \leq \epsilon$.

$|\theta^* - \hat{\theta}| \le \epsilon$.

Solution:

Recall that by definition,

$$\mathrm{TV}\left(\mathbf{P}_{\hat{\theta}}, \mathbf{P}_{\theta^*}\right) = \max_{A \subset E} \left|\mathbf{P}_{\hat{\theta}}(A) - \mathbf{P}_{\theta^*}(A)\right|$$

where the maximum is over all events A. Since we are given that $\mathrm{TV}\left(\mathbf{P}_{\hat{\theta}}, \mathbf{P}_{\theta^*}\right) \le \epsilon$, we conclude that $\left|\mathbf{P}_{\hat{\theta}}(A) - \mathbf{P}_{\theta^*}(A)\right| \le \epsilon$ for every event A. Hence, the first choice is correct.

Let A be the event given by the interval (a, b). Then,

$$\left|\mathbf{P}_{\theta^*}(a \le X \le b) - \mathbf{P}_{\hat{\theta}}(a \le Y \le b)\right| \le \epsilon$$

is the same as saying $\left|\mathbf{P}_{\hat{\theta}}(A) - \mathbf{P}_{\theta^*}(A)\right| \le \epsilon$. Thus, the second choice is true as well.

The third choice, "$|\theta^* - \hat{\theta}| \le \epsilon$.", is incorrect. In general, even if distributions $\mathbf{P}_{\hat{\theta}}$ and $\mathbf{P}_{\theta^*}$ are close, there is no reason to expect the parameters θ^* and $\hat{\theta}$ to be close. To conclude that the estimated parameter is close to the true parameter given their distributions are close, we would need some assumptions on the map $\theta \mapsto \mathbf{P}_\theta$. No such assumption is given here.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

讨论

显示讨论

主题： Unit 3 Methods of Estimation:Lecture 8: Distance measures between distributions / 4. Introduction to Total Variation Distance

认证证书是什么?