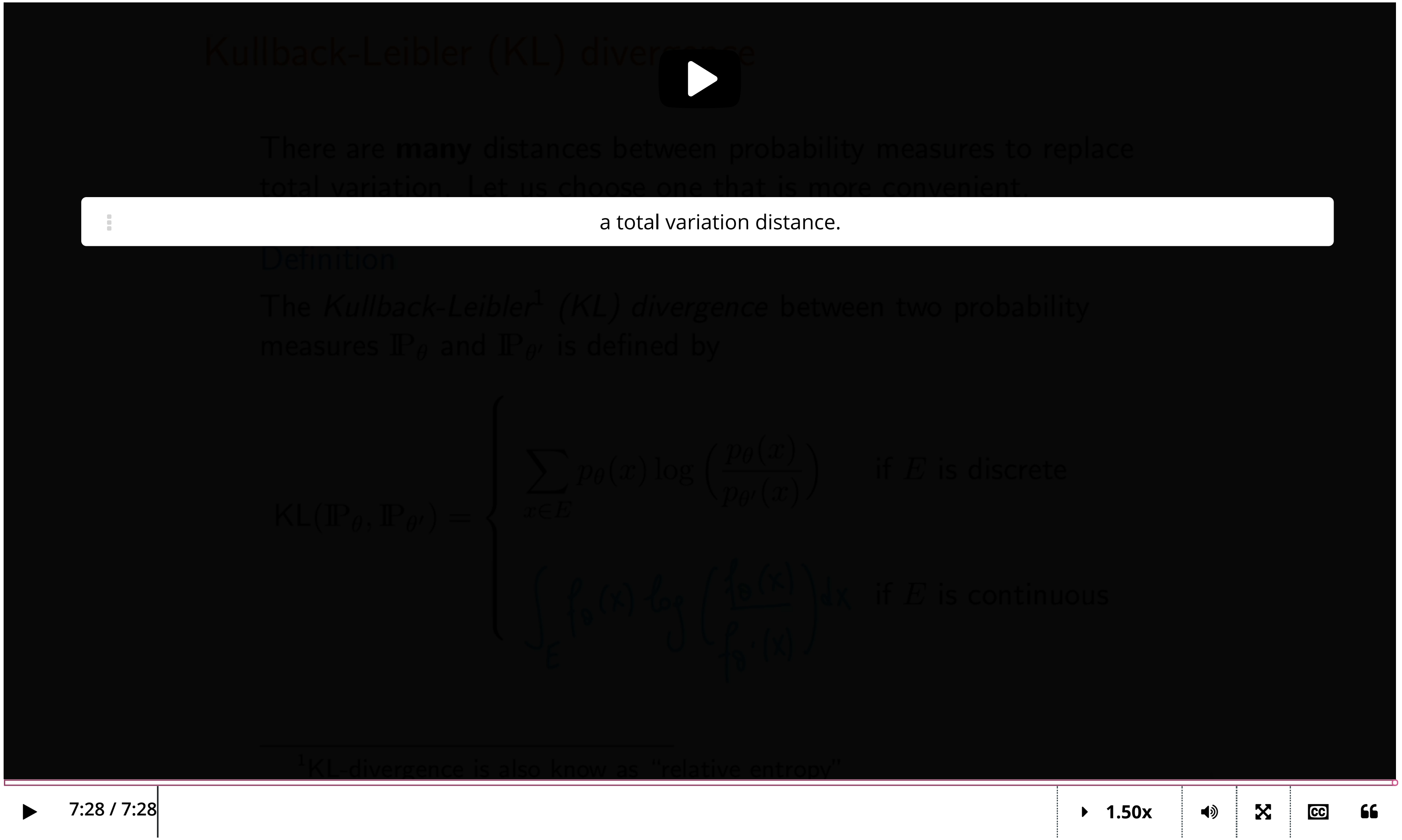
# 10. Motivation and Introduction to the Kullback-Leibler (KL) Divergence

## An Estimation Strategy and Definition of Kullback-Leibler (KL) Divergence




### Video
Download video file

### Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

---

**Definition of Kullback-Leibler (KL) Divergence**

Let $\mathbf{P}$ and $\mathbf{Q}$ be **discrete** probability distributions with pmfs $p$ and $q$ respectively. Let's also assume $\mathbf{P}$ and $\mathbf{Q}$ have a common sample space $E$. Then the **KL divergence** (also known as **relative entropy** ) between $\mathbf{P}$ and $\mathbf{Q}$ is defined by

$$\mathrm{KL}\left(\mathbf{P},\mathbf{Q}\right)=\sum_{x\in E}p\left(x\right)\ln\left(\frac{p\left(x\right)}{q\left(x\right)}\right),$$

where the sum is only over the support of $\mathbf{P}$.

**Why do we sum only over the support of P?**

Show

Analogously, if $\mathbf{P}$ and $\mathbf{Q}$ are **continuous** probability distributions with pdfs $p$ and $q$ on a common sample space $E$, then

$$\mathrm{KL}\left(\mathbf{P},\mathbf{Q}\right)=\int_{x\in E}p\left(x\right)\ln\left(\frac{p\left(x\right)}{q\left(x\right)}\right)dx,$$

where the integral is again only over the support of $\mathbf{P}$.

## Computing KL Divergence I

1/1 point (graded)
Let $X\sim\mathbf{P}_X=\mathrm{Ber}\left(1/2\right)$ and let $Y\sim\mathbf{P}_Y=\mathrm{Ber}\left(1/2\right)$. What is $\mathrm{KL}\left(\mathbf{P}_X,\mathbf{P}_Y\right)$?

$\mathrm{KL}\left(\mathbf{P}_X,\mathbf{P}_Y\right)=$ 0 ✔ **Answer:** 0.0

**Solution:**

Let $p$ be the pmf of the distribution $\mathrm{Ber}\left(1/2\right)$. Note that the sample space is the discrete set $E=\{0,1\}$. Then

$$\begin{aligned} \mathrm{KL}\left(\mathbf{P}_X, \mathbf{P}_Y\right) &= p\left(1\right) \ln\left(p\left(1\right) / p\left(1\right)\right) + p\left(0\right) \ln\left(p\left(0\right) / p\left(0\right)\right) \\ &= \left(1/2\right) \ln\left(1\right) + \left(1/2\right) \ln\left(1\right) = 0. \end{aligned}$$

**Remark:** Although KL divergence is not a distance on probability distributions (as we defined above), it does satisfy some of the axioms. For example,

- $\mathrm{KL}\left(\mathbf{P}, \mathbf{Q}\right) \geq 0$ (nonnegative), and
- $\mathrm{KL}\left(\mathbf{P}, \mathbf{Q}\right) = 0$ only if $\mathbf{P}$ and $\mathbf{Q}$ are the same distribution (definite).

Note that the result of this problem is consistent with the second property.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

## Discussion

Hide Discussion

**Topic:** Unit 3 Methods of Estimation:Lecture 8: Distance measures between distributions / 10. Motivation and Introduction to the Kullback-Leibler (KL) Divergence

Add a Post

Show all posts ▼ by recent activity ▼

? what's the KL divergence for a discrete distirbution and a continuous distribution?
just as the title, thank you. 5