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 <u>Privacy Policy</u>.





Lecture 8: Distance measures

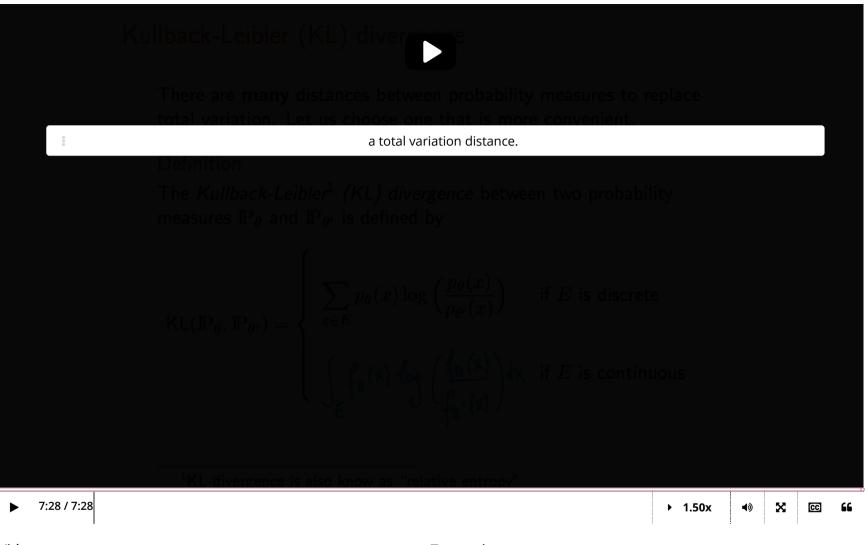
<u>Course</u> > <u>Unit 3 Methods of Estimation</u> > <u>between distributions</u>

10. Motivation and Introduction to

> the Kullback-Leibler (KL) Divergence

Currently enrolled in **Audit Track** (expires December 25, 2019) <u>Upgrade (\$300)</u>

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

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

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

**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

$$ext{KL}\left(\mathbf{P},\mathbf{Q}
ight) = \sum_{x \in E} p\left(x
ight) \ln \left(rac{p\left(x
ight)}{q\left(x
ight)}
ight),$$

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

Why do we sum only over the support of P?

<u>Show</u>

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

$$ext{KL}\left(\mathbf{P},\mathbf{Q}
ight) = \int_{x \in E} p\left(x
ight) \ln \left(rac{p\left(x
ight)}{q\left(x
ight)}
ight) dx,$$

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

## Computing KL Divergence I

1/1 point (graded)

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

$$\mathrm{KL}\left(\mathbf{P}_{X},\mathbf{P}_{Y}
ight)=egin{array}{c} \mathsf{0} \end{array}$$
 Answer: 0.0

## **Solution:**

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

$$egin{aligned} \mathrm{KL}\left(\mathbf{P}_{X},\mathbf{P}_{Y}
ight) &= p\left(1
ight)\ln\left(p\left(1
ight)/p\left(1
ight)
ight) + p\left(0
ight)\ln\left(p\left(0
ight)/p\left(0
ight)
ight) \\ &= \left(1/2
ight)\ln\left(1
ight) + \left(1/2
ight)\ln\left(1
ight) = 0. \end{aligned}$$

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

- $\mathrm{KL}\left(\mathbf{P},\mathbf{Q}
  ight)\geq0$  (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 ▼

what's the KL divergence for a discrete distirbution and a continuous distribution?

just as the title, thank you.

by recent activity ▼

style="color: blue;">

yes

style="color: blue;">

yes

style="color: blue;">

style="color: bl

## **Audit Access Expires Dec 24, 2019**

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