

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](#).



[Course](#) > [Unit 3 Methods of Estimation](#) > [Lecture 8: Distance measures between distributions](#)

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

Currently enrolled in **Audit Track** (expires December 25, 2019) [Upgrade \(\\$300\)](#)

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

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

Kullback-Leibler (KL) divergence

There are many distances between probability measures to replace total variation. Let us choose one that is more convenient: a total variation distance.

Definition

The Kullback-Leibler¹ (KL) divergence between two probability measures \mathbb{P}_θ and $\mathbb{P}_{\theta'}$ is defined by

$$\text{KL}(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \begin{cases} \sum_{x \in E} p_\theta(x) \log \left(\frac{p_\theta(x)}{p_{\theta'}(x)} \right) & \text{if } E \text{ is discrete} \\ \int_E p_\theta(x) \log \left(\frac{p_\theta(x)}{p_{\theta'}(x)} \right) d\chi & \text{if } E \text{ is continuous} \end{cases}$$

¹KL-divergence is also known as "relative entropy"

▶ 7:28 / 7:28

▶ 1.50x

🔊

🔍

📄

🗨️

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

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

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

Why do we sum only over the support of \mathbf{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

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

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

Computing KL Divergence I

1/1 point (graded)

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

$\text{KL}(\mathbf{P}_X, \mathbf{P}_Y) =$

✓ Answer: 0.0

Solution:

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

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

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

- $\text{KL}(\mathbf{P}, \mathbf{Q}) \geq 0$ (nonnegative), and
- $\text{KL}(\mathbf{P}, \mathbf{Q}) = 0$ 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

i 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 distribution and a continuous distribution?](#)
just as the title, thank you.

5

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**

Learn About Verified Certificates

© All Rights Reserved