

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](#)

11. Properties of the Kullback-Leibler (KL) Divergence

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

11. Properties of the Kullback-Leibler (KL) Divergence

Properties of Kullback-Leibler (KL) Divergence



(Caption will be displayed when you start playing the video.)



Video

[Download video file](#)

Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

Computing KL Divergence II

3/3 points (graded)

Let $X \sim \mathbf{P}_X = \text{Ber}(1/2)$ and let $Y \sim \mathbf{P}_Y = \text{Ber}(1/3)$. What is $\text{KL}(\mathbf{P}_X, \mathbf{P}_Y)$?

(If applicable, enter $\ln(\mathbf{x})$ for $\ln(x)$.)

$\text{KL}(\mathbf{P}_X, \mathbf{P}_Y) =$ ✓ Answer: 0.0588915

What is $\text{KL}(\mathbf{P}_Y, \mathbf{P}_X)$?

$\text{KL}(\mathbf{P}_Y, \mathbf{P}_X) =$ ✓ Answer: 0.05663301

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

☐ Yes

☒ No



STANDARD NOTATION

Solution:

Let f and g denote the pmfs of $\text{Ber}(1/2)$ and $\text{Ber}(1/3)$, respectively. Note that the sample space is $E = \{0, 1\}$. Then

$$\begin{aligned}\text{KL}(\mathbf{P}_X, \mathbf{P}_Y) &= \sum_{x \in \{0,1\}} f(x) \ln(f(x)/g(x)) \\ &= (1/2) \ln(3/2) + (1/2) \ln(3/4) \approx 0.0588915\end{aligned}$$

Next,

$$\begin{aligned}\text{KL}(\mathbf{P}_Y, \mathbf{P}_X) &= \sum_{x \in \{0,1\}} g(x) \ln(g(x)/f(x)) \\ &= (1/3) \ln(2/3) + (2/3) \ln(4/3) \approx 0.05663301\end{aligned}$$

Remark: In general, we have the formula

$$\text{KL}(\text{Ber}(p), \text{Ber}(q)) = p \ln \left(\frac{p}{q} \right) + (1-p) \ln \left(\frac{1-p}{1-q} \right).$$

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Properties of KL Divergence I

2/2 points (graded)

Let \mathbf{P} be a distribution such that $\text{KL}(\text{Ber}(1/2), \mathbf{P}) = 0$. What can we conclude about \mathbf{P} ?

☒ $\mathbf{P} = \text{Ber}(1/2)$.

☐ It is possible that $\mathbf{P} = \text{Ber}(p)$ for any $0 \leq p \leq 1$.

☐ \mathbf{P} could be any Gaussian distribution with mean 0 and variance $1/4$.

☐ None of the above.



What property of the KL divergence did you use to make your conclusion?

☐ Symmetric

☐ Nonnegative

☒ Definite☐ Triangle Inequality**Solution:**

The definite property of the KL divergence implies that if $\text{KL}(\mathbf{P}, \mathbf{Q}) = 0$, then \mathbf{P} and \mathbf{Q} are the same distribution. Hence, we use this property to conclude that $\mathbf{P} = \text{Ber}(1/2)$.

Note that while the KL divergence is nonnegative and definite, it is not a distance because it does not satisfy the triangle inequality nor is it symmetric.

[Submit](#)

You have used 1 of 2 attempts

i Answers are displayed within the problem

(Optional) Why does the KL divergence take only non-negative values?

[Show](#)

Discussion

[Hide Discussion](#)

Topic: Unit 3 Methods of Estimation:Lecture 8: Distance measures between distributions / 11. Properties of the Kullback-Leibler (KL) Divergence

[Add a Post](#)

Show all posts ▼

by recent activity ▼



Staff: For moment, I thought I had all answers

Start: For a moment, I thought I lost all progress...

but apparently the link to [lecture 4][1] in the optional proof section above takes to an archived version of the course, you might want to take it down. [1]: <https://courses.edx.org/...>

2

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