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

Properties of Kullback-Leibler (KL) Divergence

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

0:00 / 0:00 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Computing KL Divergence II

3/3 points (graded)

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

(If applicable, enter **ln(x)** for $\ln\,(x)$.)

$\mathrm{KL}\left(\mathbf{P}_X, \mathbf{P}_Y\right) =$ 0.05889151782819174 ✓ **Answer:** 0.0588915

What is $\mathrm{KL}\left(\mathbf{P}_Y, \mathbf{P}_X\right)$?

$\mathrm{KL}\left(\mathbf{P}_Y, \mathbf{P}_X\right) =$ 0.056633012265132426 ✓ **Answer:** 0.05663301

Is $\mathrm{KL}\left(\mathbf{P}_X, \mathbf{P}_Y\right) = \mathrm{KL}\left(\mathbf{P}_Y, \mathbf{P}_X\right)$?

○ Yes

● No

✓

STANDARD NOTATION

Solution:

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

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

Next,

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

Remark: In general, we have the formula

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

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 $\mathrm{KL}\left(\mathrm{Ber}\left(1/2\right), \mathbf{P}\right) = 0$. What can we conclude about $\mathbf{P}$?

- ● $\mathbf{P} = \mathrm{Ber}\left(1/2\right)$.
- ○ It is possible that $\mathbf{P} = \mathrm{Ber}\left(p\right)$ 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 $\mathrm{KL}\left(\mathbf{P},\mathbf{Q}\right)=0$, then $\mathbf{P}$ and $\mathbf{Q}$ are the same distribution. Hence, we use this property to conclude that $\mathbf{P}=\mathrm{Ber}\left(1/2\right)$.

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

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 a moment, I thought I lost all progress

Staff: 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.... 2