



Bookmarks

► Introduction

▼ 1. Probability and Inference

Introduction to Probability (Week 1)

Exercises due Sep 22, 2016 at 02:30 IST



Probability Spaces and Events (Week 1)

Exercises due Sep 22, 2016 at 02:30 IST



Random Variables (Week 1)

Exercises due Sep 22, 2016 at 02:30 IST



Jointly Distributed Random Variables (Week 2)

Exercises due Sep 29, 2016 at 02:30 IST



Conditioning on Events (Week 2)

Exercises due Sep 29, 2016 at 02:30 IST



1. Probability and Inference &gt; Measuring Randomness (Week 4) &gt; Exercise: Mutual Information



Bookmark

## Exercise: Mutual Information

(2/2 points)

Consider the following joint probability table for random variables  $X$  and  $Y$ . We'll compute the mutual information  $I(X; Y)$  of random variables  $X$  and  $Y$  step-by-step.

		$Y$		
		0	1	2
$X$	0	0.10	0.09	0.11
	1	0.08	0.07	0.07
	2	0.18	0.13	0.17

Mutual information is about comparing the joint distribution of  $X$  and  $Y$  with what the joint distribution would be if  $X$  and  $Y$  were actually independent.

**Homework 1 (Week 2)**

Homework due Sep 29, 2016 at 02:30 IST

**Inference with Bayes' Theorem for Random Variables (Week 3)**

Exercises due Oct 06, 2016 at 02:30 IST

**Independence Structure (Week 3)**

Exercises due Oct 06, 2016 at 02:30 IST

**Homework 2 (Week 3)**

Homework due Oct 06, 2016 at 02:30 IST

**Notation Summary (Up Through Week 3)****Mini-project 1: Movie Recommendations (Weeks 3 and 4)**

Mini-projects due Oct 13, 2016 at 02:30 IST

**Decisions and Expectations (Week 4)**

Exercises due Oct 13, 2016 at 02:30 IST

**Measuring Randomness (Week 4)**

Exercises due Oct 13, 2016 at 02:30 IST



In Python (where we won't explicitly store the labels of the rows and columns):

```
import numpy as np
joint_prob_XY = np.array([[0.10, 0.09, 0.11], [0.08, 0.07, 0.07], [0.18, 0.13, 0.17]])
```

The marginal distributions  $p_X$  and  $p_Y$  are given by:

```
prob_X = joint_prob_XY.sum(axis=1)
prob_Y = joint_prob_XY.sum(axis=0)
```

Next, we produce what the joint probability table would be if  $X$  and  $Y$  were actually independent:

```
joint_prob_XY_indep = np.outer(prob_X, prob_Y)
```

At this point, we have the joint distribution of  $X$  and  $Y$  (denoted  $p_{X,Y}$ ) stored in code as `joint_prob_XY`, and also what the joint distribution would be if  $X$  and  $Y$  were independent (denoted  $p_X p_Y$ ) stored in code as `joint_prob_XY_indep`. The mutual information of  $X$  and  $Y$  is precisely given by the KL divergence between  $p_{X,Y}$  and  $p_X p_Y$ :

$$I(X; Y) = D(p_{X,Y} \parallel p_X p_Y) = \sum_x \sum_y p_{X,Y}(x, y) \log_2 \frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)}.$$

- What is  $I(X; Y)$ ? Provide just the number and don't write "bits" at the end. We suggest that you code a Python function that computes the information divergence between any two distributions, and then you can just plug in `joint_prob_XY` and `joint_prob_XY_indep`.

(Please be precise with at least **5** decimal places, unless of course the answer doesn't need that many decimal places. You could also put a fraction.)

## Towards Infinity in Modeling Uncertainty (Week 4)

Exercises due Oct 13, 2016 at 02:30 IST



## Homework 3 (Week 4)

Homework due Oct 13, 2016 at 02:30 IST



0.00226108299607



Answer: 0.0022610829960697087

- Are  $X$  and  $Y$  independent?

☐ Yes

☒ No 

### Solution:

- What is  $I(X; Y)$ ? Provide just the number and don't write "bits" at the end.

**Solution:** With NumPy, we can code the information divergence as a one-liner:

```
info_divergence = lambda p, q: np.sum(p * np.log2(p / q))
```

If you haven't seen `lambda` before, the above single line is equivalent to:

```
def info_divergence(p, q):  
    return np.sum(p * np.log2(p / q))
```

Then, to compute  $I(X; Y)$ , we do:

```
mutual_info_XY = info_divergence(joint_prob_XY, joint_prob_XY_indep)
```

Printing out `mutual_info_XY` yields **0.0022610829960697087** bits.

- Are  $X$  and  $Y$  independent?

**Solution:** The answer is **no**. If  $X$  and  $Y$  were independent, then  $p_{X,Y}$  and  $p_X p_Y$  would be the same distribution, which means that the KL divergence between them would be 0, which means that the mutual information  $I(X; Y)$  would be 0. But as we just computed,  $I(X; Y)$  is nonzero!

*You have used 3 of 5 submissions*

© All Rights Reserved



© 2016 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY  
OPENedX®

