MITx 6.419x
Data Analysis: Statistical Modeling and Computation in Applications

Help HuitianDiao

Course Progress Dates Discussion Resources

Course / Module 5: Environmental Data and G... / Sensing and Analyzing global patter...

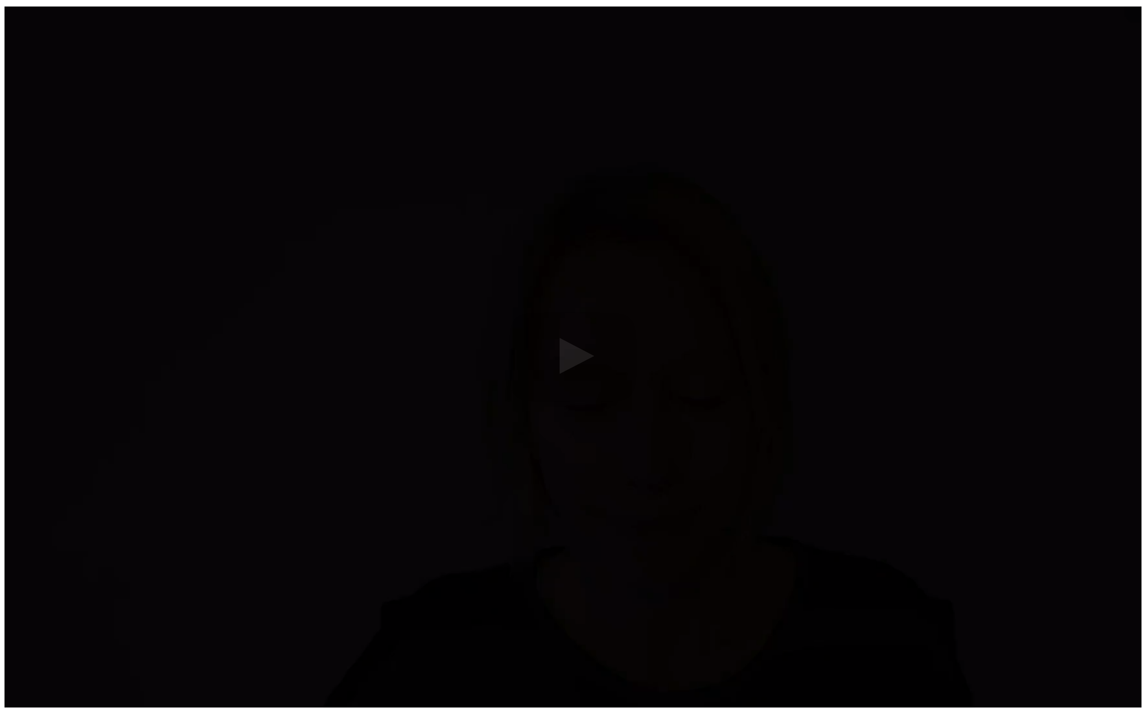
Previous Next

6. Beyond Linear Dependencies

Bookmark this page

Exercises due May 21, 2021 19:59 EDT

Beyond Linear Dependencies



22:42 / 22:42 1.50x

are significant.

And so once you can check it in the network sense, the patterns

and phenomena that you observe, would you also

observe them with a random graph with a degree sequence,

for example?

The second thing is, you can look at whether it's actually

a function of the time series.

And you can scramble the time series values

and see whether your observations still holds.

So are they actually a phenomena that's induced by the data,

or would a random scramble also give you that?

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

So far, we have used correlation as a measure of connection between two random variables. However, correlation is a linear property, and in general, one might have other non-linear connections or dependencies on the data that will not be captured by measuring correlations.

For example, consider two random variables Z and Y. The random variable Z can be $+1$ or -1 with equal probability. On the other hand, the random variable Y can be 1 or 2 with equal probability. Now, define a new random variable $X = Y * Z$

Correlation of 2 Variables

1 point possible (graded)

Consider the two random variables defined above. Are X and Y **correlated** (in terms of the correlation coefficient)?

- Yes
- No

No

Submit You have used 0 of 1 attempt

Independence of 2 Variables

1 point possible (graded)

Consider the two random variables defined above. Are X and Y independent.

Yes

No

Submit You have used 0 of 1 attempt

In the two exercises above, we have seen that dependency and correlation do not imply each other. In order to capture other forms of dependency, we introduce the concept of **mutual information**.

Definition 6.1 The mutual information between two random variables X and Y is defined as

$$I(X,Y) = H(X) - H(X \mid Y),$$

where $H(X)$ is the (discrete) entropy of the random variable X taking values in $\mathcal{X}$, defined as

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x)),$$

The entropy of a discrete distribution is a non-negative quantity, and is bounded above by the equivalent entropy for a uniform distribution. Thus, for a discrete distribution over n values:

$$0 \leq H(X) \leq \log n.$$

$H(X,Y)$ is the (discrete) joint entropy of the random variable X and Y taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log(p(x,y)) \leq H(X) + H(Y),$$

and $H(X|Y)$ is the (discrete) conditional entropy of the random variable X and Y taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, defined as

$$H(X|Y) = H(X,Y) - H(Y).$$

Note that $H(X) \geq H(X|Y)$.

The base of the logarithm in the above formulae depends on the units we wish to measure information in. A base of 2 is common in computer science, and measures the information in **bits**. In the physical sciences, a natural logarithm is often used and measures the information in units of **nats**.

The property of interest for mutual information is that $I(X,Y) = 0$ if and only if X and Y are independent.

Entropy of a uniform distribution.

1 point possible (graded)
Let $p(x) = 0.25$ for $x \in \mathcal{X} = \{1, 2, 3, 4\}$ such that p is a uniform distribution.

We now wish to ask what the entropy of this distribution is in **bits**. To do this, we need to compute the entropy using the base 2 logarithm:

$$H_2(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2(p(x)),$$

What is the entropy in **bits**?

$H_2(X) =$

Submit You have used 0 of 3 attempts

Entropy of a joint distribution.

1 point possible (graded)
Let $p(x, y)$ be a joint distribution over $x \in \{0, 1\}$ and $y \in \{0, 1\}$ such that

$$\begin{aligned} p(0,0) &= 0.5 \\ p(0,1) &= 0 \\ p(1,0) &= 0.25 \\ p(1,1) &= 0.25. \end{aligned}$$

What is the entropy in **bits**? Note that

$$\lim_{z \to 0} z \log z \to 0$$

$H_2(X, Y) =$

Submit You have used 0 of 3 attempts

Conditional entropy.

1 point possible (graded)
Let $p(x, y)$ be a joint distribution as above:

$$\begin{aligned} p(0,0) &= 0.5 \\ p(0,1) &= 0 \\ p(1,0) &= 0.25 \\ p(1,1) &= 0.25. \end{aligned}$$

What is the conditional entropy of X given Y in **bits**? Please provide your answer to **three significant figures** (graded to 1% tolerance).

$H_2(X|Y) =$

Submit You have used 0 of 3 attempts

Mututal information.

1 point possible (graded)
Let us return to the example of two random variables Z and Y. The random variable Z can be $+1$ or -1 with equal probability. On the other hand, the random variable Y can be 1 or 2 with equal probability. Now, define a new random variable $X = Y * Z$

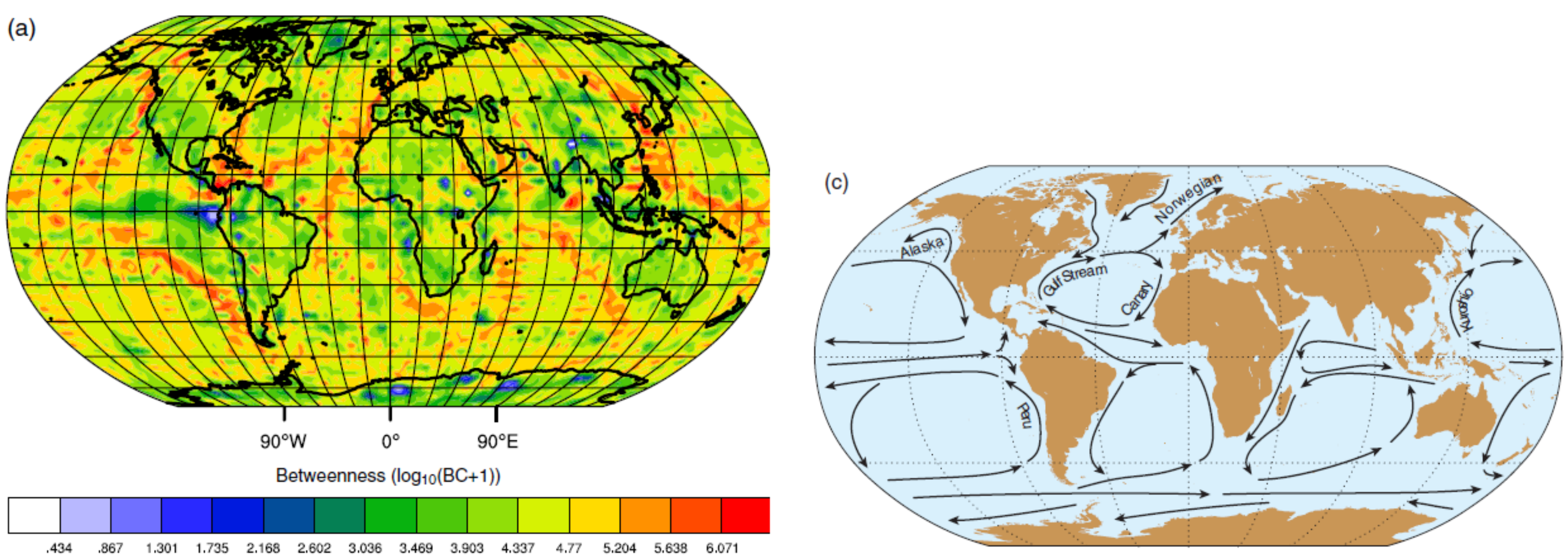
What is the mutual information of X and Y in **bits**?

$I_2(X, Y) =$

Submit You have used 0 of 3 attempts

The main idea with the introduction of mutual information to quantify non-linear relations among random variables. Thus, following the same procedure as above, one can define edges in a graph if two nodes have mutual information higher than some predefined threshold.

The figure below shows the betweenness centrality for a network that was built based on mutual information between the temperature measurements on the map. Red values indicate a higher centrality, while blue values indicate lower centrality in the generated network. One can observe that the zones with high centrality approximately match the zones where there are some ocean surface currents, as shown in the map.



50: Betweenness centrality for a network built from mutual information relations.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Sensing and Analyzing global patterns of dependence / 6. Beyond Linear Dependencies

Add a Post

Show all posts by recent activity

? Conditional entropy
The value of H(Y) in answer is right? A little different with the fomula of H(X). 1

[Staff] Minor typo in answer to mutual information question
In the answer to the mutual information question, the values for y in the p(x,y) are by error mentioned to be +1 and -1, whereas the c... 1

< Previous Next >

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store