

Joint entropy

In information theory, **joint entropy** is a measure of the uncertainty associated with a set of variables.^[2]

Contents

Definition

Properties

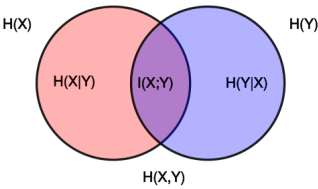
- Nonnegativity
- Greater than individual entropies
- Less than or equal to the sum of individual entropies

Relations to other entropy measures

Joint differential entropy

- Definition
- Properties

References



A misleading^[1] Venn diagram showing additive, and subtractive relationships between various information measures associated with correlated variables X and Y. The area contained by both circles is the joint entropy H(X,Y). The circle on the left (red and violet) is the individual entropy H(X), with the red being the conditional entropy H(X|Y). The circle on the right (blue and violet) is H(Y), with the blue being H(Y|X). The violet is the mutual information I(X;Y).

Definition

The joint Shannon entropy (in bits) of two discrete random variables **X** and **Y** with images **X** and **Y** is defined as^{[3]:16}

$$H(X,Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x,y) \log_2 [P(x,y)]$$

(Eq.1)

where **x** and **y** are particular values of **X** and **Y**, respectively, **P(x,y)** is the joint probability of these values occurring together, and **P(x,y) log₂[P(x,y)]** is defined to be 0 if **P(x,y) = 0**.

For more than two random variables **X₁, ..., X_n** this expands to

$$H(X_1, \dots, X_n) = - \sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_n \in \mathcal{X}_n} P(x_1, \dots, x_n) \log_2 [P(x_1, \dots, x_n)]$$

(Eq.2)

where **x₁, ..., x_n** are particular values of **X₁, ..., X_n**, respectively, **P(x₁, ..., x_n)** is the probability of these values occurring together, and **P(x₁, ..., x_n) log₂[P(x₁, ..., x_n)]** is defined to be 0 if **P(x₁, ..., x_n) = 0**.

Properties

Nonnegativity

The joint entropy of a set of random variables is a nonnegative number.

$$H(X,Y) \geq 0$$

$$H(X_1, \dots, X_n) \geq 0$$

Greater than individual entropies

The joint entropy of a set of variables is greater than or equal to the maximum of all of the individual entropies of the variables in the set.

$$H(X,Y) \geq \max [H(X), H(Y)]$$

$$H(X_1, \dots, X_n) \geq \max_{1 \leq i \leq n} \left\{ H(X_i) \right\}$$

Less than or equal to the sum of individual entropies

The joint entropy of a set of variables is less than or equal to the sum of the individual entropies of the variables in the set. This is an example of subadditivity. This inequality is an equality if and only if **X** and **Y** are statistically independent.^{[3]:30}

$$H(X,Y) \leq H(X) + H(Y)$$

$$H(X_1, \dots, X_n) \leq H(X_1) + \dots + H(X_n)$$

Relations to other entropy measures

Joint entropy is used in the definition of conditional entropy.^{[3]:22}

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

$$H(X_1, \ldots, X_n) = \sum_{k=1}^n H(X_k | X_{k-1}, \ldots, X_1)$$

It is also used in the definition of mutual information^{[3]:21}

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

In quantum information theory, the joint entropy is generalized into the joint quantum entropy.

Joint differential entropy

Definition

The above definition is for discrete random variables and no more valid in the case of continuous random variables. The continuous version of discrete joint entropy is called *joint differential (or continuous) entropy*. Let **X** and **Y** be a continuous random variables with a joint probability density function **f(x, y)**. The differential joint entropy **h(X, Y)** is defined as^{[3]:249}

$$h(X, Y) = - \int_{\mathcal{X}, \mathcal{Y}} f(x, y) \log f(x, y) \, dx dy$$
(Eq.3)

For more than two continuous random variables **X**₁, ..., **X**_n the definition is generalized to:

$$h(X_1, \ldots, X_n) = - \int f(x_1, \ldots, x_n) \log f(x_1, \ldots, x_n) \, dx_1 \ldots dx_n$$
(Eq.4)

The integral is taken over the support of **f**. It is possible that the integral does not exist in which case we say that the differential entropy is not defined.

Properties

As in the discrete case the joint differential entropy of a set of random variables is smaller or equal than the sum of the entropies of the individual random variables:

$$h(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^n h(X_i)$$
^{[3]:253}

The following chain rule holds for two random variables:

$$h(X, Y) = h(X|Y) + h(Y)$$

In the case of more than two random variables this generalizes to:^{[3]:253}

$$h(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n h(X_i | X_1, X_2, \ldots, X_{i-1})$$

Joint differential entropy is also used in the definition of the mutual information between continuous random variables:

$$I(X, Y) = h(X) + h(Y) - h(X, Y)$$

References

1. D.J.C. Mackay. *Information theory, inferences, and learning algorithms*.^{:141}

2. Theresa M. Korn; Korn, Granino Arthur. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Publications. ISBN 0-486-41147-8.

3. Thomas M. Cover; Joy A. Thomas. *Elements of Information Theory*. Hoboken, New Jersey: Wiley. ISBN 0-471-24195-4.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Joint_entropy&oldid=924098872"

This page was last edited on 1 November 2019, at 20:07 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.