

# Conditional mutual information

In probability theory, particularly information theory, the **conditional mutual information**<sup>[1][2]</sup> is, in its most basic form, the expected value of the mutual information of two random variables given the value of a third.

## Contents

Definition

In terms of pmf's for discrete distributions

In terms of pdf's for continuous distributions

Some identities

More general definition

Note on notation

Properties

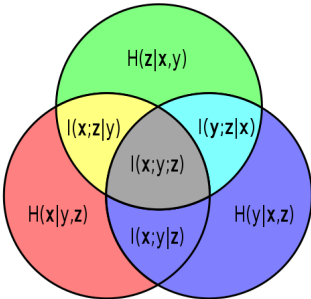
Nonnegativity

Interaction information

Chain rule for mutual information

Multivariate mutual information

References



Venn diagram of information theoretic measures for three variables  $x$ ,  $y$ , and  $z$ , represented by the lower left, lower right, and upper circles, respectively. The conditional mutual informations  $I(x;z|y)$ ,  $I(y;z|x)$  and  $I(x;y|z)$  are represented by the yellow, cyan, and magenta regions, respectively.

## Definition

For random variables  $X$ ,  $Y$ , and  $Z$  with support sets  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$ , we define the conditional mutual information as

$$I(X;Y|Z) = \int_{\mathcal{Z}} D_{\text{KL}}(P_{(X,Y)|Z} \| P_{X|Z} \otimes P_{Y|Z}) dP_Z$$

This may be written in terms of the expectation operator:  $I(X;Y|Z) = \mathbb{E}_Z[D_{\text{KL}}(P_{(X,Y)|Z} \| P_{X|Z} \otimes P_{Y|Z})]$ .

Thus  $I(X;Y|Z)$  is the expected (with respect to  $Z$ ) Kullback–Leibler divergence from the conditional joint distribution  $P_{(X,Y)|Z}$  to the product of the conditional marginals  $P_{X|Z}$  and  $P_{Y|Z}$ . Compare with the definition of mutual information.

## In terms of pmf's for discrete distributions

For discrete random variables  $X$ ,  $Y$ , and  $Z$  with support sets  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$ , the conditional mutual information  $I(X;Y|Z)$  is as follows

$$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} p_Z(z) \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{X,Y|Z}(x,y|z) \log \frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)}$$

where the marginal, joint, and/or conditional probability mass functions are denoted by  $p$  with the appropriate subscript. This can be simplified as

$$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{X,Y,Z}(x,y,z) \log \frac{p_Z(z)p_{X,Y,Z}(x,y,z)}{p_{X,Z}(x,z)p_{Y,Z}(y,z)}.$$

## In terms of pdf's for continuous distributions

For (absolutely) continuous random variables  $X$ ,  $Y$ , and  $Z$  with support sets  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$ , the conditional mutual information  $I(X;Y|Z)$  is as follows

$$I(X;Y|Z) = \int_{\mathcal{Z}} \left( \int_{\mathcal{Y}} \int_{\mathcal{X}} \log \left( \frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)} \right) p_{X,Y|Z}(x,y|z) dx dy \right) p_Z(z) dz$$

where the marginal, joint, and/or conditional probability density functions are denoted by  $p$  with the appropriate subscript. This can be simplified as

$$I(X;Y|Z) = \int_{\mathcal{Z}} \int_{\mathcal{Y}} \int_{\mathcal{X}} \log \left( \frac{p_Z(z)p_{X,Y,Z}(x,y,z)}{p_{X,Z}(x,z)p_{Y,Z}(y,z)} \right) p_{X,Y,Z}(x,y,z) dx dy dz.$$

## Some identities

Alternatively, we may write in terms of joint and conditional entropies as<sup>[3]</sup>

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

This can be rewritten to show its relationship to mutual information

$$I(X;Y|Z) = I(X;Y,Z) - I(X;Z)$$

usually rearranged as **the chain rule for mutual information**

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

Another equivalent form of the above is<sup>[4]</sup>

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

Like mutual information, conditional mutual information can be expressed as a Kullback–Leibler divergence:

$$I(X;Y|Z) = D_{\text{KL}}[p(X,Y,Z)||p(X|Z)p(Y|Z)p(Z)].$$

Or as an expected value of simpler Kullback–Leibler divergences:

$$\begin{aligned} I(X;Y|Z) &= \sum_{z \in \mathcal{Z}} p(Z=z) D_{\text{KL}}[p(X,Y|z)||p(X|z)p(Y|z)], \\ I(X;Y|Z) &= \sum_{y \in \mathcal{Y}} p(Y=y) D_{\text{KL}}[p(X,Z|y)||p(X|Z)p(Z|y)]. \end{aligned}$$

## More general definition

A more general definition of conditional mutual information, applicable to random variables with continuous or other arbitrary distributions, will depend on the concept of **regular conditional probability**. (See also.<sup>[5][6]</sup>)

Let  $(\Omega, \mathcal{F}, \mathfrak{P})$  be a probability space, and let the random variables  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$  each be defined as a Borel-measurable function from  $\Omega$  to some state space endowed with a topological structure.

Consider the Borel measure (on the  $\sigma$ -algebra generated by the open sets) in the state space of each random variable defined by assigning each Borel set the  $\mathfrak{P}$ -measure of its preimage in  $\mathcal{F}$ . This is called the pushforward measure  $\mathbf{X}_*\mathfrak{P} = \mathfrak{P}(\mathbf{X}^{-1}(\cdot))$ . The **support of a random variable** is defined to be the topological support of this measure, i.e. **supp  $\mathbf{X}$  = supp  $\mathbf{X}_*\mathfrak{P}$** .

Now we can formally define the conditional probability measure given the value of one (or, via the product topology, more) of the random variables. Let  $\mathbf{M}$  be a measurable subset of  $\Omega$ , (i.e.  $\mathbf{M} \in \mathcal{F}$ ), and let  $\mathbf{x} \in \text{supp } \mathbf{X}$ . Then, using the disintegration theorem:

$$\mathfrak{P}(\mathbf{M}|\mathbf{X}=\mathbf{x}) = \lim_{U \ni \mathbf{x}} \frac{\mathfrak{P}(\mathbf{M} \cap \{\mathbf{X} \in U\})}{\mathfrak{P}(\{\mathbf{X} \in U\})} \qquad \text{and} \qquad \mathfrak{P}(\mathbf{M}|\mathbf{X}) = \int_{\mathbf{M}} d\mathfrak{P}(\omega|\mathbf{X}=\mathbf{X}(\omega)),$$

where the limit is taken over the open neighborhoods  $U$  of  $\mathbf{x}$ , as they are allowed to become arbitrarily smaller with respect to set inclusion.

Finally we can define the conditional mutual information via Lebesgue integration:

$$I(X;Y|Z) = \int_{\Omega} \log \left( \frac{d\mathfrak{P}(\omega|\mathbf{X},\mathbf{Z})}{d\mathfrak{P}(\omega|\mathbf{Z})} \frac{d\mathfrak{P}(\omega|\mathbf{Y},\mathbf{Z})}{d\mathfrak{P}(\omega|\mathbf{X},\mathbf{Y},\mathbf{Z})} \right) d\mathfrak{P}(\omega),$$

where the integrand is the logarithm of a Radon–Nikodym derivative involving some of the conditional probability measures we have just defined.

## Note on notation

In an expression such as  $I(\mathbf{A};\mathbf{B}|\mathbf{C})$ ,  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  need not necessarily be restricted to representing individual random variables, but could also represent the joint distribution of any collection of random variables defined on the same probability space. As is common in probability theory, we may use the comma to denote such a joint distribution, e.g.  $I(\mathbf{A}_0,\mathbf{A}_1;\mathbf{B}_1,\mathbf{B}_2,\mathbf{B}_3|\mathbf{C}_0,\mathbf{C}_1)$ . Hence the use of the semicolon (or occasionally a colon or even a wedge  $\wedge$ ) to separate the principal arguments of the mutual information symbol. (No such distinction is necessary in the symbol for joint entropy, since the joint entropy of any number of random variables is the same as the entropy of their joint distribution.)

## Properties

### Nonnegativity

It is always true that

$$I(X;Y|Z) \geq 0,$$

for discrete, jointly distributed random variables  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$ . This result has been used as a basic building block for proving other inequalities in information theory, in particular, those known as Shannon-type inequalities.

### Interaction information

Conditioning on a third random variable may either increase or decrease the mutual information: that is, the difference  $I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ , called the interaction information, may be positive, negative, or zero. This is the case even when random variables are pairwise independent. Such is the case when:

$$X \sim \text{Bernoulli}(0.5), Z \sim \text{Bernoulli}(0.5), \quad Y = \begin{cases} X & \text{if } Z = 0 \\ 1 - X & \text{if } Z = 1 \end{cases}$$

in which case  $X, Y$  and  $Z$  are pairwise independent and in particular  $I(X; Y) = 0$ , but  $I(X; Y|Z) = 1$ .

Chain rule for mutual information

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z)$$

Multivariate mutual information

The conditional mutual information can be used to inductively define a **multivariate mutual information** in a set- or measure-theoretic sense in the context of information diagrams. In this sense we define the multivariate mutual information as follows:

$$I(X_1; \dots; X_{n+1}) = I(X_1; \dots; X_n) - I(X_1; \dots; X_n|X_{n+1}),$$

where

$$I(X_1; \dots; X_n|X_{n+1}) = \mathbb{E}_{X_{n+1}} [D_{\text{KL}}(P_{(X_1, \dots, X_n)|X_{n+1}} \| P_{X_1|X_{n+1}} \otimes \dots \otimes P_{X_n|X_{n+1}})].$$

This definition is identical to that of interaction information except for a change in sign in the case of an odd number of random variables. A complication is that this multivariate mutual information (as well as the interaction information) can be positive, negative, or zero, which makes this quantity difficult to interpret intuitively. In fact, for  $n$  random variables, there are  $2^n - 1$  degrees of freedom for how they might be correlated in an information-theoretic sense, corresponding to each non-empty subset of these variables. These degrees of freedom are bounded by various Shannon- and non-Shannon-type inequalities in information theory.

References

1. Wyner, A. D. (1978). "A definition of conditional mutual information for arbitrary ensembles". *Information and Control*. **38** (1): 51–59. doi:10.1016/s0019-9958(78)90026-8 (<https://doi.org/10.1016%2Fs0019-9958%2878%2990026-8>).

2. Dobrushin, R. L. (1959). "General formulation of Shannon's main theorem in information theory". *Uspekhi Mat. Nauk*. **14**: 3–104.

3. Cover, Thomas; Thomas, Joy A. (2006). *Elements of Information Theory* (2nd ed.). New York: Wiley-Interscience. ISBN 0-471-24195-4.

4. Decomposition on Math.StackExchange (<https://math.stackexchange.com/q/1863993>)

5. Regular Conditional Probability (<https://planetmath.org/encyclopedia/ConditionalProbabilityMeasure.html>) on PlanetMath (<https://planetmath.org/>)

6. D. Leao, Jr. et al. *Regular conditional probability, disintegration of probability and Radon spaces*. *Proyecciones*. Vol. 23, No. 1, pp. 15–29, May 2004, Universidad Católica del Norte, Antofagasta, Chile PDF (<http://www.scielo.cl/pdf/proy/v23n1/art02.pdf>)

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Conditional\\_mutual\\_information&oldid=950992956](https://en.wikipedia.org/w/index.php?title=Conditional_mutual_information&oldid=950992956)"

This page was last edited on 14 April 2020, at 22:51 (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.