Author date: 2020-01-09

Extending Shannon entropy to continuous variables

E. Castedo Ellerman (castedo@castedo.com)

Copyright:
creativecommons.org/licenses/by/4.0/
2020 © The Authors. This document is distributed under a Creative Commons Attribution 4.0 International license.

What qualifies as a satisfactory extension of Shannon entropy to continuous variables is somewhat subjective. What extension is desirable is highly dependent on what one wishes to measure.

Shannon defines a continuous entropy [1] but clarifies

> *"In the discrete case the entropy measures in an absolute way the randomness of the chance variable. In the continuous case the measurement is relative to the coordinate system."*

The continuous entropy defined by Shannon can be negative. The type of extension considered in this document is one that measures entropy "in an absolute way".

Statistical Variance

Information is also a measure of uncertainty. A common, if not the most common, measure of uncertainty for continuous variables is statistical variance. Like entropy, variance shares the property that the measure for a combination of two independent sources is the sum of the measures for the respective sources. More formally, given any two independent variables X and Y

$$\begin{array}{rclcl} \mathrm{H}((X,Y)) & = & \mathrm{H}(X,Y) & = & \mathrm{H}(X)+\mathrm{H}(Y) \\ \mathrm{Var}((X,Y)) & = & \mathrm{Var}(X+Y) & = & \mathrm{Var}(X)+\mathrm{Var}(Y) \end{array}$$

A preference in entropy extension could be to match variance for continuous variables. This document considers the class of entropy extensions that align with variance for continuous variables.

Mutual Information

A measurement derived from entropy is

$$I(X;Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}((X,Y))$$

which Shannon refers to as the *actual* rate of information transmission [1] (where X and Y are the start and end of a noisy communication channel). More recent authors refer to this as *mutual information* [2]. One of many interpretations is that $I(X;Y)$ measures the amount of information about X provided by Y.

A notable property of mutual information (for two variables) is that it is zero if and only if the two random variables are independent.

Variance Explained

Variance explained is another measurement which also captures a sense of how much information one variable provides about another. Formally, given random variables X and Y, variance explained is $\mathrm{Var}(\mathrm{E}(X|Y))$ using the definition of conditional expectation [3] where $\mathrm{E}(X|Y)$ is a random variable.

The variance explained by an independent variable is zero. However, unlike mutual information, zero variance explained does not imply independence. Consider a random variable X that takes the values $\{-1, 0, 1\}$ with equal probability. The random variable $Y = |X|$ explains zero variance, but X and Y are not independent. Intuitively, the random variable Y does provide some information about X. It informs whether X is zero or not. In this sense, something analogous to mutual information is a more appropriate measure of how much information Y provides about X.

Random Objects vs Variables

A *random object* is a function with a domain of a probability space [3]. When the function values are real numbers, it is a *random variable* (or *real random object*). A *finite random object* means a random object that takes on finitely many values. Or in other words, the range of a *finite random object* is a set of finite size.

Both entropy and variance are functions of random objects. In the case of Shannon entropy, the random object is finite (with values often referred to as symbols). In the case of variance, the random object is a random variable (possibly a vector of real numbers in $\mathbb{R}^n$). The distances between values of a finite random variable affect variance, but not Shannon entropy.

Desirable Extension Properties

Some random variables are also finite random objects, but their variance and Shannon entropy are not necessarily equal. Any extension must have some extra input beyond just a random object to determine whether the output is statistical variance vs Shannon entropy. Let G represent a desirable extension with G_h and G_v denoting the cases when some extra input determines Shannon entropy vs statistical variance, respectively.

Extending mutual information to continuous variables is a desirable property. Since entropy is equal to the mutual information between a variable and itself, describing an extension of mutual information, will also describe an extension of entropy.

For finite random objects, the Shannon entropy case should satisfy

$$G_h(X, Y) = I(X;Y)$$
$$G_h(X, X) = \mathrm{H}(X)$$

Similarly, for real random objects (random variables), the variance case should satisfy

$$G_v(X, X) = \mathrm{Var}(X)$$

Lastly, the notable property to satisfy is mutual information extended to continuous variables:

$$G_v(X, Y) = 0 \text{ if and only if } X \text{ and } Y \text{ are independent}$$

References

1. Shannon CE, Weaver W. The mathematical theory of communication. Urbana: Univ. of Illinois Press; 1998.
2. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2006.
3. Ash RB, Doléans-Dade C, Ash RB. Probability and measure theory. 2nd ed. San Diego: Harcourt/Academic Press; 2000.