Jensen–Shannon divergence

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **Jensen–Shannon divergence** is a popular method of measuring the similarity between two probability distributions. It is also known as **information radius (IRad)**[1] or **total divergence to the average**.[2] It is based on the Kullback–Leibler divergence, with some notable (and useful) differences, including that it is symmetric and it is always a finite value. The square root of the Jensen–Shannon divergence is a metric.[3][4]

Contents

Definition

Consider the set $M_+^1(A)$ of probability distributions where A is a set provided with some σ-algebra of measurable subsets. In particular we can take A to be a finite or countable set with all subsets being measurable.

The Jensen–Shannon divergence (JSD) $M_+^1(A) \times M_+^1(A) \to [0, \infty)$ is a symmetrized and smoothed version of the Kullback–Leibler divergence $D(P \parallel Q)$. It is defined by

$$JSD(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M)$$

where $M = \frac{1}{2}(P + Q)$

If A is countable, a more general definition, allowing for the comparison of more than two distributions, is:

$$JSD(P_1, P_2, \ldots, P_n) = H\left(\sum_{i=1}^{n} \pi_i P_i\right) - \sum_{i=1}^{n} \pi_i H(P_i)$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are the weights for the probability distributions $P_1, P_2, \ldots, P_n$ and $H(P)$ is the Shannon entropy for distribution P. For the two-distribution case described above,

$$P_1 = P, P_2 = Q, \pi_1 = \pi_2 = \frac{1}{2}.$$

Bounds

The Jensen–Shannon divergence is bounded by 1, given that one uses the base 2 logarithm.[5]

$$0 \leq JSD(P \parallel Q) \leq 1$$

For log base e, or ln, which is commonly used in statistical thermodynamics, the upper bound is ln(2):

$$0 \leq JSD(P \parallel Q) \leq ln(2)$$

Relation to mutual information

Jensen–Shannon divergence is the mutual information between a random variable X from a mixture distribution $M = \frac{P+Q}{2}$ and a binary indicator variable Z where $Z = 1$ if X is from P and $Z = 0$ if X is from Q.

$$\begin{aligned} I(X;Z) &= H(X) - H(X|Z) \\ &= -\sum M \log M + \frac{1}{2}\left[\sum P \log P + \sum Q \log Q\right] \\ &= -\sum \frac{P}{2} \log M - \sum \frac{Q}{2} \log M + \frac{1}{2}\left[\sum P \log P + \sum Q \log Q\right] \\ &= \frac{1}{2}\sum P\left(\log P - \log M\right) + \frac{1}{2}\sum Q\left(\log Q - \log M\right) \\ &= JSD(P \parallel Q) \end{aligned}$$

It follows from the above result that Jensen–Shannon divergence is bounded by 0 and 1 because mutual information is non-negative and bounded by $H(Z) = 1$.

One can apply the same principle to the joint and product of marginal distribution (in analogy to Kullback–Leibler divergence and mutual information) and to measure how reliably one can decide if a given response comes from the joint distribution or the product distribution—given that these are the only possibilities.[6]

Quantum Jensen–Shannon divergence

The generalization of probability distributions on density matrices allows to define quantum Jensen–Shannon divergence (QJSD).[7][8] It is defined for a set of density matrices $(\rho_1, \ldots, \rho_n)$ and probability distribution $\pi = (\pi_1, \ldots, \pi_n)$ as

$$QJSD(\rho_1, \ldots, \rho_n) = S\left(\sum_{i=1}^{n} \pi_i \rho_i\right) - \sum_{i=1}^{n} \pi_i S(\rho_i)$$

where $S(\rho)$ is the von Neumann entropy. This quantity was introduced in quantum information theory, where it is called the Holevo information: it gives the upper bound for amount of classical information encoded by the quantum states $(\rho_1, \ldots, \rho_n)$ under the prior distribution π (see Holevo's theorem)[9] Quantum Jensen–Shannon divergence for $\pi = (\frac{1}{2}, \frac{1}{2})$ and two density matrices is a symmetric function, everywhere defined, bounded and equal to zero only if two density matrices are the same. It is a square of a metric for pure states[10] but it is unknown whether the metric property holds in general.[8] The Bures metric is closely related to the quantum JS divergence; it is the quantum analog of the Fisher information metric.

Applications

The Jensen–Shannon divergence has been applied in bioinformatics and genome comparison,[11][12] and in protein surface comparison.[13]

Notes

1. ^ Hinrich Schütze; Christopher D. Manning (1999). *Foundations of Statistical Natural Language Processing* (http://nlp.stanford.edu/fsnlp/). Cambridge, Mass: MIT Press. p. 304. ISBN 0-262-13360-1.
2. ^ Dagan, Ido; Lillian Lee, Fernando Pereira (1997). "Similarity-Based Methods For Word Sense Disambiguation" (http://citeseer.ist.psu.edu/dagan97similaritybased.html). *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*: pp. 56–63. Retrieved 2008-03-09.
3. ^ Endres, D. M.; J. E. Schindelin (2003). "A new metric for probability distributions". *IEEE Trans. Inf. Theory* **49** (7): pp. 1858–1860. doi:10.1109/TIT.2003.813506 (http://dx.doi.org/10.1109%2FTIT.2003.813506).
4. ^ Ôsterreicher, F.; I. Vajda (2003). "A new class of metric divergences on probability spaces and its statistical applications". *Ann. Inst. Statist. Math.* **55** (3): pp. 639–653. doi:10.1007/BF02517812 (http://dx.doi.org/10.1007%2FBF02517812).
5. ^ Lin, J. (1991). "Divergence measures based on the shannon entropy" (https://www.cise.ufl.edu/~anand/sp06/jensen-shannon.pdf). *IEEE Transactions on Information Theory* **37** (1): 145–151. doi:10.1109/18.61115 (http://dx.doi.org/10.1109%2F18.61115).
6. ^ Schneidman, Elad; Bialek, W; Berry, M.J. 2nd (2003). "Synergy, Redundancy, and Independence in Population Codes". *Journal of Neuroscience* **23** (37): 11539–11553. PMID 14684857 (//www.ncbi.nlm.nih.gov/pubmed/14684857).
7. ^ A. P. Majtey, P. W. Lamberti, and D. P. Prato (2005) "Jensen–Shannon divergence as a measure of distinguishability between mixed quantum states," Phys. Rev. A 72, 052310
8. ^ ***a b*** J. Briet, P. Harremoes (2009) "Properties of classical and quantum Jensen–Shannon divergence", *Phys. Rev. A*, 79, 052311 doi:10.1103/PhysRevA.79.052311 (http://dx.doi.org/10.1103%2FPhysRevA.79.052311)
9. ^ A.S.Holevo (1973) "Bounds for the quantity of information transmitted by a quantum communication channel", *Problemy Peredachi Informatsii* 9, 3–11 (1973) (in Russian) (English translation: *Probl. Inf. Transm.*, 9, 177–183 (1975)) MR456936 (http://www.ams.org/mathscinet-getitem?mr=456936))
10. ^ S. L. Braunstein, C. M. Caves (1994) "Statistical distance and the geometry of quantum states", *Phys. Rev. Lett.*, 72 3439 doi:10.1103/PhysRevLett.72.3439 (http://dx.doi.org/10.1103%2FPhysRevLett.72.3439)
11. ^ G.E. Sims, S.R. Jun, G. Wu. & S.H. Kim (2009) "Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions" (http://www.ncbi.nlm.nih.gov/pubmed/19188606), *Proc. Natl. Acad. Sci. USA*. 106(8):2677–82 PubMed (http://www.ncbi.nlm.nih.gov/pubmed/19188606)
12. ^ S. Itzkovitz, E. Hodis, E. Segal (2010) "Overlapping codes within protein-coding sequences," *Genome Res.*, 20:1582–1589
13. ^ Y. Ofran & B. Rost (2003). "Analysing Six Types of Protein–Protein Interfaces"

(http://citeseer.ist.psu.edu/ofran03analysing.html), *J. Mol. Biol.*, 325: 377—38

Further reading

- Bent Fuglede and Flemming Topsøe "Jensen-Shannon Divergence and Hilbert space embedding" (http://www.math.ku.dk/~topsoe/ISIT2004JSD.pdf), University of Copenhagen, Department of Mathematics
- F. Nielsen (2011) "A family of statistical symmetric divergences based on Jensen's inequality",arXiv:1009.4004

External links

- Ruby gem for calculating JS divergence (https://github.com/evansenter/diverge)

Retrieved from "http://en.wikipedia.org/w/index.php?title=Jensen–Shannon_divergence&oldid=555507042"
Categories: Statistical distance measures
