

# Universal Estimation of Information Measures

Sergio Verdú

Dept. of Electrical Engineering

Princeton University

Princeton, NJ 08544

**Abstract**—In this presentation I will give an overview of the state of the art in universal estimation of:

- Entropy
- Divergence
- Mutual Information

with emphasis on recent algorithms we have proposed with H. Cai, S. Kulkarni and Q. Wang. These algorithms converge to the desired quantities without any knowledge of the statistical properties of the observed data, under several conditions such as stationary-ergodicity in the case of discrete processes, and memorylessness in the case of analog data. A sampling of the literature in this topic is given below.

- [17] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov, A. J. Wyner. "Nonparametric entropy estimation for stationary processes and random fields, with applications to english text". *IEEE Trans. Inform. Theory*, Vol. 44, No. 3, pp 1319-1327, May 1998.
- [18] A. N. Quas. "An entropy estimation for a class of infinite alphabet processes". *Theory Probab. Appl.*, Vol. 43, pp 496-507, 1995.

## REFERENCES

- [1] C. E. Shannon. "Prediction and Entropy of Printed English". *Bell System Technical Journal*, pp 50-64, 1951.
- [2] H. Cai, S. Kulkarni, S. Verdu., Universal estimation of entropy and divergence via block sorting, Proc. on IEEE ISIT 2002, pp 433, Lausanne, Switzerland, July 2002.
- [3] H. Cai, S. Kulkarni, S. Verdu, Universal entropy estimation via block sorting, *IEEE Trans. Inform. Theory*, Vol. 50, No. 7, pp 1551-1561, July 2004.
- [4] P. J. Moreno, P. P. Ho and N. Vasconcelos, "A Kullback-Leibler Divergence Based Kernel for SVM Classification in Multimedia Applications", HPL-2004-4, HP Laboratories Cambridge.
- [5] A. Kaltchenko, Algorithms for Estimation of Information Distance with Application to Bioinformatics and Linguistics," 2004 Canadian Conference on Electrical and Computer Engineering, Niagara Falls, Canada, May 2004.
- [6] P. Hall, "On Kullback-Leibler Loss and Density Estimation", *The Annals of Statistics*, Vol. 15, No. 4, pp 1491-1519, Dec. 1987.
- [7] J. Beirlant, E. J. Dudewicz, L. Györfi and E. C. van der Meulen, "Non-parametric Entropy Estimation: An Overview", *International Journal of Mathematical and Statistical Sciences*. 6 (1) : pp 17-39, 1997.
- [8] J. Ziv, N. Merhav, "A Measure of Relative Entropy Between Individual Sequences with Application to Universal Classification", *IEEE Transactions on Information Theory*, Vol. 39, No. 4, pp 1270-1279, July 1993.
- [9] A. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of Entropic Spanning Graphs", *Signal Processing Magazine, IEEE*, Vol. 19, No. 5, pp 85-95, Sept. 2002.
- [10] G. A. Darbellay and I. Vajda, "Estimation of the Information by an Adaptive Partitioning of the Observation Space", *IEEE Transactions On Information Theory*, Vol. 45, No. 4, pp 1315 - 1321, May 1999.
- [11] G. Lugosi and A. Nobel, "Consistency of Data-driven Histogram Methods for Density Estimation and Classification", *The Annals of Statistics*, Vol. 24, No. 2, pp 687-706, 1996.
- [12] I. Gijbels and J. Mielniczuk, "Asymptotic Properties of Kernel Estimators of the Radon-Nikodym Derivative with Applications to Discriminant Analysis", *Statistica Sinica*, pp 261-278, 1995.
- [13] L. Paninski, "Estimation of Entropy and Mutual information" *Neural Computation*, Vol. 15, No. 6, pp 1191-1253, June 2003.
- [14] S. Panzeri and A. Treves, "Analytical Estimates of Limited Sampling Biases in Different Information Measures", *Network: Computation in Neural Systems*, pp 87-107, 1996.
- [15] D. Benedetto, E. Caglioti, V. Loreto. "Language Trees and Zipping". *Physical Review Letters*, Vol. 88, No. 4, 28 Jan, 2002.
- [16] M. Gutman. "Asymptotically optimal classification for multiple tests with empirically observed statistics". *IEEE Trans. Inform. Theory*, Vol. 35, No. 2, pp 401-408, March 1989.