

Information Theory and Statistics

Tom Cover¹

Departments of Electrical Engineering and Statistics, Stanford University, Durand 121, Stanford, CA 94305-4055, USA,
email: cover@isl.stanford.edu

Abstract — The main theorems in information theory and statistics are put in context, the differences are discussed, and some of the open research problems are mentioned.

I. INTRODUCTION

Probability theory has produced a number of strong general statements — truths about stochastic processes that give random processes a deterministic flavor. These successes include the strong law of large numbers, the central limit theorem, the law of the iterated logarithm, the ergodic theorem, and limit theorems for Markov processes.

Information theory, on the other hand, has been primarily motivated by an attempt to optimize certain processes, for example, to minimize the description length of random processes or to maximize the number of distinguishable signals in the presence of noise. This different orientation — optimization — has led to a number of additional insights which contribute to the body of knowledge in probability theory. For example, the central limit theorem can be proved by way of the entropy power inequality, yielding a monotonic convergence to the limit. And the law of large numbers has a counterpart in the asymptotic equipartition property, which says that all ergodic stochastic processes can be considered as a uniform distribution over a small set of typical sequences characterized by the entropy rate.

II. SPECIFICS

We will demonstrate some of the points of intersection of information theory and statistics, and mention some problems in physics and computer science that require a rigorous probabilistic treatment.

The discussion will include the following:

1. Chernoff information, error exponents, large deviation theory.
2. The geometry of information.
3. Structure of ergodic processes, the AEP and the Slepian Wolf theorem.
4. The common basis for the Cramer-Rao, entropy power, Brunn-Minkowski, and Heisenberg uncertainty inequalities. (See Dembo.)
5. Entropy rate (compressibility limits), channel capacity (distinguishability limits). The duality of the two.
6. The central limit theorem and the entropy power inequality. (See Barron.)
7. Information loss and the second law. The argument that entropy will be lost when mass is thrown into a black hole, together with the even stronger belief that entropy increases (the second law of thermodynamics), led Beckenstein and Hawking to argue that the mass of

the black hole (which increases when matter is thrown into it) is proportional to its entropy (the logarithm of the number of ways in which it could be made) thus preserving the second law.

8. Entropy increase. The \dot{H} theorem in statistical mechanics shows that entropy increases with time. But the laws of physics are time reversible. What is going on?
9. Investment processes. Duality with data compression.

III. REMARKS

Certain theorems from information theory like the asymptotic equipartition property (the Shannon-MacMillan-Breiman theorem) may deserve to be considered part of the hard core of probability theory. Yet other results in information theory like the entropy power inequality turn out to play an important role in interpreting the central limit theorem. And finally, some of the tools in information theory may have important roles to play in physics, just as ergodic theory, developed in the 1930s, resolved some of the problems in statistical mechanics.

ACKNOWLEDGEMENTS

The work of Imre Csiszar has greatly influenced my understanding.

REFERENCES

- [1] I. Csiszar, T. Cover and B.S. Choi. Conditional Limit Theorems under Markov Conditioning. *IEEE Transactions on Information Theory*, IT-33(6): 788-801, November 1987.
- [2] A. Dembo, T. Cover and J. Thomas. Information Theoretic Inequalities. *IEEE Transactions on Information Theory*, 37(6):1501-1518, November 1991.
- [3] A. Barron. Entropy and the Central Limit Theorem. *The Annals of Probability*, 14(1):336-342, 1986.

¹This work was partially supported by NSF Grant NCR-9205663 and JSEP Contract DAAH04-94-G-0058.