On the role of Information Theory in Probability and Statistics

FLEMMING TOPSØE[1]

University of Copenhagen, Department of Mathematics, Universitetsparken 5, 2100 Copenhagen, Denmark
topsoe@math.ku.dk

AMS classification: 94 A 15, 62 B 10, 91 A 80
Keywords: Information Theory, Statistics, Game Theory.

Information theory is becoming more and more important for several sciences including probability theory and, especially, statistics. The development can be traced in works of Kullback [1], Čencov [2], Csiszár [3], [4], Dembo and Zeituni [5] and others. General textbooks are appearing (Jessop (1995), Applebaum (1996)). More is to come and exciting research is in the pipeline.

There are two reasons for the successes. Firstly, the theory works with "the right concepts" as evidenced in Shannon [6] and in Kullback and Leibler [7] where the most important quantity for applications to statistics, *information divergence*, is introduced. Note that divergence, in contrast to entropy, also applies to continuous distributions.

Interpretations take the human expressive power and ability of efficient descriptions into consideration, i.e. *complexity* considerations are essential. Entropy is a case of minimum description length, and divergence is a similar quantity which, more realistically, takes into account that often one will not know the underlying distribution, but may still want to measure performance in relation to an ideal situation with a known statistic.

The second reason behind the successes of information theory in statistics is the ability to model *conflict situations*, typically via two-person-zero sum games. Suffice it here to point to the maximum entropy principle and the minimum information discrimination principle. The proper treatment of these principles involves two players, say the statistician and "the system", responsible for the choice of the "actual model". Following Jaynes (1957), we realize that what the statistician is really up against is not so much an immaterial person as "the system", "God" or whatever, but more so his own *knowledge* about the system under study.

The presentation will clarify some of these points, based partly on ongoing, still unpublished research of Peter Harremoës which builds on works of Csiszár ([3], [4]), the author ([8]) and others.

References

[1] S. Kullback, *Informaton Theory and Statistics.* New York: Wiley, 1959.

[2] N. N. Čencov, *Statistical Decision Rules and Optimal Decision.* Moscow: Nauka, 1972. In Russian.

[3] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probab.*, vol. 3, pp. 146–158, 1975.

[4] I. Csiszár, "Sanov property, generalized I-projection and a conditional limit theorem," *Ann. Probab.*, vol. 12, pp. 768–793, 1984.

[5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications.* Boston: Jones and Bartlett Publishers International, 1993.

[6] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.

[7] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[8] F. Topsøe, "Information theoretical optimization techniques," *Kybernetika*, vol. 15, pp. 8–27, 1979.

[1]Supported by The Danish Natural Science Research Council and by INTAS (project 00-738)