Information Theory has been primarily motivated by problems in telecommunication systems, e.g.,
minimizing the description length of random processes or maximizing the number of distinguishable
signals in the presence of noise. This perspective — optimization — has led to a number of insights
that contribute to the body of knowledge in Statistics and Probability Theory. This course will
discuss the applications of information theoretic methods in statistics.

The course will begin with introductory information theory. Notions of entropy, mutual informa-
tion, and Kullback-Leibler divergence will be covered. Their significance in data compression and
error correcting codes will be brought out via the source and channel coding theorems.

Three topics that will be addressed are

- **Information Geometry**: Notions of $I$-projections on sets of probability measures will be
covered. Examples will be drawn from problems encountered in maximum entropy and maxi-
mum likelihood estimation, the EM algorithm, etc. Projection of a probability mass function
onto a linear family of probability mass functions, and iterative algorithms for finding the
minimum divergence between two convex sets of probability mass functions will be studied.

- **Large Deviations**: Sanov’s theorem and related results will be presented from an infor-
mation theoretic viewpoint. The role of $I$-divergence will be highlighted, and applications of
large deviations principles will be illustrated via the evaluation of error exponents for selected
estimation problems.

- **Redundancy and Data Compression**: Lossy and lossless compression of data will be
investigated. In particular, Rissanen’s Minimum Description Length (MDL) principle, and
other model based techniques for universal lossless data compression will be studied, and
connections with statistical modeling will be explored.

Participants are expected to be familiar with probability theory at the level of an advanced under-
graduate or core graduate course. Exposure to graduate level information theory is useful, but is
not a prerequisite.

**Course Organization**: The material will be presented in a style somewhere between a regular
course a graduate seminar. Sections from assorted text books will be presented as preparatory
material, followed by discussion of select papers from technical journals.
Assessment: Grades will be based on in-class participation (33%) and a take-home examination (67%) given in the latter half of the course.

References

[1] Information Theory Notes, From courses taught by I. Csiszár (1989) and P. Narayan (1994) at the University of Maryland, and P. Shields (1996) at the University of Toledo.


