Mokshay M. Madiman

Ph.D. Dissertation Summary

Thesis Title: Topics in Information Theory, Probability and Statistics

Advisor: Ioannis Kontoyiannis

Submitted to: Division of Applied Mathematics, Brown University

Status: Defended successfully August 2005

This dissertation explores three topics at the intersection of information theory, probability and statistics. The first, titled "The Minimum Description Length Principle in Lossy Data Compression", develops a theoretical framework for lossy source coding that treats it as a statistical problem, in analogy to the approach to universal lossless coding suggested by Rissanen's Minimum Description Length (MDL) principle. Two methods for selecting efficient compression algorithms are proposed, based on lossy variants of the Maximum Likelihood and MDL principles. Their theoretical performance is analyzed, and it is shown under appropriate assumptions that the MDL approach to universal lossy coding identifies the optimal model class of lossy codes.

The second part, titled "Measure Concentration for Compound Poisson Distributions", considers the phenomenon of measure concentration for Lipschitz functions, of non-negative integer valued (i.e., $\mathbb{Z}_+$-valued) compound Poisson random variables. New concentration inequalities with power law decay are proved when the Lévy measure has finite $(1+\delta)^{th}$ moment for some $\delta > 0$. These are among the first examples of polynomial concentration in the literature. In addition to a new modification of the Herbst argument, a new, elementary proof of a logarithmic Sobolev inequality for compound Poisson distributions is presented.

The third part, titled "An information-theoretic approach to Compound Poisson Approximation" develops an information-theoretic approach to obtaining compound Poisson approximation bounds for sums of $\mathbb{Z}_+$-valued random variables. First, a general and simple relative entropy bound for approximating an arbitrary function of possibly dependent $\mathbb{Z}_+$-valued random variables using independent compound Poissons is proved. This yields a simple approximation bound for sums. In order to obtain refined bounds, three useful "local information quantities" are defined and analyzed. In particular, one of these quantities is used to obtain optimal-order total variation bounds for the case of "nearly homogeneous portfolios".