Optimal Universal Coding and Density Estimation

Yu, Bin

University of Wisconsin
1210 W. Dayton Street
Department of Statistics
Madison, WI 53706

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

Approved for public release; distribution unlimited.

Research progress has been made in the areas of empirical processes for mixing sequences, information theory, minimax estimation theory in source coding and non-parametric statistics, and Markov chain Monte Carlo (MCMC) methods.

Rates of convergence and Central Limit Theorems results have been obtained for empirical processes of dependent data, and they are very useful for studying statistical models with dependence structure. On the important MCMC convergence diagnostic problem, regeneration points have been introduced into the Markov chain using the split-chain technique; so has been a global approach based on the estimated $L^1$ error and the Cusum path plot. Making connections between information theory and statistics, we obtained an information-theoretic result on the rate of convergence of a D-semifaithful code, and we also introduced non-parametric minimax lower bound techniques into bounding from below the redundancy in source coding.

Density estimation, Empirical process, Markov chain Monte Carlo, Minimax, Minimum Description Length, Rate-distortion theory.

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED

UL
Final Report on ARO Grant (DAAL03-91-G-0107)
awarded to University of Wisconsin-Madison

Principle Investigator: Bin Yu
Current Mailing Address:
Statistics Department
367 Evans Hall
University of California
Berkeley, CA 94720-3860

1 Statement of Problems Studied

Under ARO grant (DAAL03-91-G-007) for the period of October, 1990 through September, 1994, the principle investigator conducted research in the areas of empirical processes for mixing sequences, information theory, minimax estimation theory in source coding and non-parametric statistics, and Markov chain Monte Carlo (MCMC) methods.

2 Summary of Research Results

With partial support of the ARO grant, seven papers were written, some in collaboration with others, on empirical processes for mixing sequences and on Markov Chain Monte Carlo methods.

Statisticians have recently turned their attention to dependent data. One way to formulate dependence in data are the mixing conditions. Rates of convergence and Central Limit Theorems results for empirical processes of dependent data are very useful for studying statistical models with dependence structure. In [1], rates of convergence results are given for empirical processes of $\beta$-mixing sequences and using the techniques developed there, optimal rates of convergence results are obtained for density estimation error in the $L^\infty$ norm and for mixing sequences in [5]. Jointly with M. Arcones at University of Utah, we provide in [2] a Central Limit Theorem of empirical processes for completely regular sequences under almost minimal conditions. In the same paper, limit theorems for U-processes are also given.

The Markov chain Monte Carlo (MCMC) methods are being studied intensively for both Bayesian and likelihood computations. The MCMC method enables us to obtain (dependent) samples from a target density from which direct sampling is difficult. Quantities of
interest of the target distribution, such as mean, variance, and tail probabilities, can then be approximated using the MCMC sample. Since the target distribution is the stationary distribution of the constructed Markov chain, the success of the MCMC methods relies crucially on our ability to assess the convergence of the chain to its equilibrium. In the joint paper [10] with P. Mykland and L. Tierney of Univ of Chicago and Univ. of Minnesota respectively, we introduce regeneration points into the Markov chain using the split-chain technique, and therefore provide a way of diagnosing the convergence of the common MCMC schemes. In [11], a global approach for convergence diagnostic is introduced based on the estimated $L^1$ error of a kernel estimator by utilizing the information contained in the unnormalized target density form. In [12] (jointly with Mykland), a simple Cusum plot is used in the MCMC context to extract more diagnostic information from a single run of MCMC. Paper [5] addresses for the first time the density estimation problem in MCMC as well. It also gives a comparison between two sampling schemes in Markov Chain Monte Carlo simulation and an assessment of the MLE based on an approximate likelihood function using the Gibbs sampler.

Information theory and Statistics have always been closely related and are now finding more common grounds when researchers from both fields trying to communicate more with each other as evidenced by the joint IEEE-IMS workshop in Virginia last month. Five paper were written on information theory related topics, and in particular Minimum Description Length (MDL) principle were to statistical problems. The joint paper [6] with T. P. Speed at UC-Berkeley contains an information-theoretic result on the rate of convergence of a D-semifaithful code, extending part of existing results in the MDL literature from noisyless codes to rate-distortion theory (or D-semifaithful codes). In [3], a well-known hypercube technique due to Assouad is used to construct a minimax lower bound on the rate of redundancy of d-th order continuous Markov sources. The lower bound is explicit enough in d that the nonexistence of a redundancy rate in the continuous case can be deduced. Paper [9] makes connections between three very useful techniques in nonparametric minimax lower bound construction. In [8], we illustrate the application of the MDL principle to a typical stochastic learning problem, where the features range over a continuum. Moreover, we show that when the object we try to learn, e.g., the probability function of the weight, lies in a parametric class, the best rate at which we can estimate it (or in other words, the best rate at which we can "learn" about it) is the same as the complexity of the model, that is, minimum description length of the model. When the model class is much larger, say a smooth nonparametric class, the "learning" rate is much slower. Paper [7] concerns the problem of assessing the performance of model selection criteria in terms of two kinds of predictions in the context of normal linear regression. The particular selection criteria considered are AIC, BIC and Predictive MDL. It illustrates using a simple model that at the heart of of the problem of model selection is still the bias and variance trade-off, and no criterion is universally better than others.
3 Publication List

References


4 Scientific Personnel Involved

Bin Yu (Principle Investigator)

Ananda Sen (Research Assistant. Ph.D. 1993 under Professor G. Bhattacharyya. Bin Yu was on the Thesis Committee.)

Yonghong Yang (Research Assistant. Ph.D. Student)

Zhen Luo (Research Assistant. Ph.D. Student)