Construction of Efficient Algorithms for the Estimation of Multivariate Probability Densities

James R. Thompson

Department of Mathematical Sciences
Rice University
Houston, Texas 77251-1892

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

Approved for public release; distribution unlimited.
CONSTRUCTION OF EFFICIENT ALGORITHMS FOR THE ESTIMATION
OF MULTIVARIATE PROBABILITY DENSITIES

JAMES R. THOMPSON

U.S. ARMY RESEARCH OFFICE

DAAG-29-82-K-0014

RICE UNIVERSITY

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

THE VIEWS, OPINIONS AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE
OF THE AUTHOR AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF
THE ARMY POSITION, POLICY OR DECISION, UNLESS SO DESIGNATED BY OTHER
DOCUMENTATION.
Statement of Problem Studied

Our task has been to develop algorithms, and, to whatever extent possible, portable software for the data-based nonparametric estimation of densities of high dimensionality. We hoped to employ graphical and interactive techniques whenever possible. In cases where the dimensionality was very high, we hoped to develop an algorithm for the finding of modes which might then be used as origins for further investigations. In addition, we hoped to examine related problem areas, such as parameter estimation for applied stochastic process models. It was hoped that interaction with the Ballistics Research Laboratory at Aberdeen would be fruitful. Also, it was hoped that the Texas Medical Center and NASA would provide useful local data sources on which to test our algorithms.
Summary of Most Important Results

The main thrust of our work has been in the nonparametric estimation and representation of densities of dimensions of three and greater. An algorithm for multicolour, time lapsed density contour display has been developed by Scott and is now portable. A 16mm film has been made of the application of this technique to estimate and display densities of dimensions three, four and five. This film has been shown at the Aberdeen Proving Ground, at the 1983 DOD Conference at the USAF Academy, at the ONR 1983 Luray Conference, in addition to some dozen nonDOD conferences. The principal data sets used in explication of the Scott algorithm are ground observations from satellite overflights.

The nonparametric estimation and representation of densities of very high dimensionality (up to one hundred) is a problem addressed in the doctoral dissertation of Steven Boswell, presently of Lincoln Laboratories. Boswell's algorithm seeks out the modes of the density which generated a data set. These modes are then to be used as origins for further analysis.

The dissertation of Hathaway uses the EM algorithm to address the problem of estimating a density as a mixture of Gaussian densities. Many of the instability problems generally encountered with mixture techniques are eliminated.

The problem of using a data base as a vehicle for simulation has been addressed by Taylor and Thompson using behind armour data from the Aberdeen Proving Ground. Their procedure gives the same effects as those which would have been obtained if the underlying density had first been estimated nonparametrically and then used to build a random number generator. However, by eliminating the density estimation step, the Taylor-Thompson algorithm provides enormous savings both in computer running times, and, more importantly, in the ease of use.

The development of simulation techniques for parameter estimation in complex stochastic process situations has been a major activity for Thompson and his colleagues at the Texas Medical Center, Atkinson, Bartoszynski and Brown. Not only are these techniques more stable and quicker than those based on approximation theoretic procedures, but, since they require only the axioms of the process, rather than some "closed form solution," they are much easier to use. A problem which had required a man year to code using standard techniques was coded in one week using the simulation based strategy. The development of this algorithm enables the use of stochastic models in real world situations which earlier had defied such applied modeling.
Publications


Scientific Personnel Supported and Ph.D. Degrees Awarded

Bartoszynski, Robert: Visiting Professor of Mathematical Sciences

Boswell, Steve: Graduate Student (presently Research Associate Lincoln Labs), Ph.D. awarded 1983.

Chiu, Shean-Tsong: Assistant Professor of Mathematical Sciences

Hathaway, Richard: Graduate Student (presently Assistant Professor Univ. of S.C.) Ph.D. awarded 1982.

Jee, Rodney: Graduate Student (hired by JPL), Ph.D. to be completed in 1985

Scott, David W.: Associate Professor of Mathematical Sciences

Stevens, Joyce: Graduate Student.

Tapia, Richard A.: Professor of Mathematical Sciences

Terrell, George R.: Visiting Assistant Professor of Mathematical Sciences

Thompson, James R.: Professor of Mathematical Sciences
END

FILMED

4-85

DTIC