THE CONSTRUCTION OF ALGORITHMS FOR THE EFFICIENT ESTIMATION OF --ETC(U)
JAN 82 J R THOMPSON

UNCLASSIFIED

ARO-15637-12-H

END

F/6 12/1
DAA029-79-C-0180
NL
The Construction of Algorithms for the Efficient Estimation of Multivariate Probability Densities

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>James R. Thompson</th>
</tr>
</thead>
</table>

### Abstract

Techniques have been proposed for estimating probability densities nonparametrically. These include penalized maximum likelihood, series, fixed kernel and nearest neighbour procedures. Special attention was given to the two and three dimensional cases. A data-based random number generator for arbitrarily high dimensions was also developed and applied to ballistics data at Aberdeen Proving Ground. A number of biomedical applications were proposed for the related problem of the nonparametric estimation of intervals using cancer data.
The Construction of Algorithms for the Efficient Estimation of Multivariate Probability Densities

Final Report

Author: James R. Thompson

January 20, 1982

U. S. Army Research Office

DAAG-29-79-C-0180
It has been a major objective of this investigation to determine means by which data from an unknown multivariate probability distribution may be most expeditiously summarized, particularly in the exploratory phase of the data analysis process. Unlike the members of exploratory data analysis (EDA) school who attempt to avoid models, our analysis is oriented toward the ultimate goal of modelling the stochastic mechanism at hand. [1] gives a survey of some of the problems outstanding in density estimation.

In [2] we give our algorithm for discrete maximum penalized likelihood estimation of a one dimensional density nonparametrically. This algorithm is now one of the most used for nonparametric probability density estimation—having been made a part of the IMSL software library (NDMPLE). It should be mentioned that this routine, as a part of the IMSL package, is on disk at practically all DOD computer centers.

In [3], three basic approaches are investigated for estimating two dimensional densities. Computer software is developed for implementing each of the approaches. We note that the two dimensional density estimation problem is actually a three dimensional problem from a representational standpoint. Accordingly, for higher dimensions simple graphing of the density is not possible.

[4] examines the nonparametric estimation of three dimensional densities. The "onion peel" algorithm is developed for this task. By this procedure, we first find the modes of the density and then use these as centers for further investigation.

A problem posed by Dr. Malcolm Taylor of the Aberdeen Proving Ground involves large scale simulations in which a multidimensional data set is to be used as a base for a pseudorandom number generator. An algorithm [5] was developed for this purpose which avoids the necessity of estimating the underlying density. It is now operational both at Aberdeen and at Rice.
[6] gives a parametric analysis of breast cancer data in an attempt to model the progress of the disease in a large population of patients. A tentative conclusion is made that in a significant proportion of cases of the disease the five year disease free equivalence of a cure does not appear valid. In [7], we discuss a model-theoretic approach for the optimal scheduling of chemotherapeutic regimes.

The technique of maximum penalized likelihood is employed in [8] for the nonparametric estimation of the intensity of metastasis formation. Contrary to classical neoplastic theory, we discovered that the appearance of new metastases is essentially constant in time.

Based on the exploratory data analysis in [8], we developed in [9] a stochastic model for the progress of cancer and used it to analyze a number of data sets from several different kinds of cancer. Our tentative conclusion is that cancer appears to proceed as a systemic disease with the throwing off of metastases relegated to a secondary function. This result gives rise to new speculation as to the causes and treatment of cancer.
REFERENCES


Personnel Supported

Chi-Chy Fwu (Ph.D. Student)
Donna Nezames (Ph.D. Student, degree granted 1980)
Richard Tapia
James R. Thompson