MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

AD-A244 721

TATION PAGE

Form Approved
OMB No. 0704-0188

2. DATE	3. REPORT TYPE AND DATES COVERED
Sept. 26, 1991	Final 1 Aug 88 - 31 Jul 91

4. TITLE AND SUBTITLE

Some Problems in Density Estimation, Modeling and Time Series Analysis

5. FUNDING NUMBERS

DAAL-03-88-K-0131

6. AUTHOR(S)

James R. Thompson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Statistics
Rice University
Houston, Texas 77251-1892

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 26108.1-MA

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Two books and 22 articles were produced under sponsorship of DAAL-03-88-K-0131. Work in density estimation resampling has been carried out with Dr. Malcolm Taylor of BRL. Simulation based estimation has been applied to problems ranging from marketing to oncology. Two doctoral dissertations were written on the subject of nonparametric regression. The effects of more active public health intervention to halt the AIDS epidemic were considered. New algorithms for density estimation in several dimensions were developed. The estimation of the parameters characterizing a time series was pursued using both frequency and time domain techniques. The possibility that chaos is an artifact of inappropriate modeling was investigated.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

92-00457

92 1 7 077

FINAL REPORT

Some Problems in Density Estimation, Modeling, and Time Series Analysis

CONTRACT OR GRANT NUMBER: DAAL-03-88-K-0131

AUTHOR OF REPORT: James R. Thompson

Department of Statistics, Rice University, Houston, Texas 77251-1892

Outline of Results

Time series analysis. Chiu created an algorithm for computing the parameters of an ARMA time series model in which the order of work is linear. Also, he developed a test for the periodic components in a time series. [1] - [5].

Ensor developed computationally efficient iterative algorithms for estimating the parameters of an autoregressive process. [6].

Nonparametric density estimation. Scott developed algorithms for estimation of densities in multivariate situations. He further developed software for the practical utilization of these algorithms by various graphical displays. Cross-validation algorithms have been an integral part of his work. [8], [10] - [15]. He is completing work on a book dealing with techniques for density estimation in which the higher dimensional case is emphasized.

Thompson has considered nonparametric density estimation as an exploratory device. He has noted how the problem of finding modes can actually become easier as the dimensionality of a data set increases, if appropriate mean update algorithms are employed. [21].

Nonparametric Regression and Hazard Function Estimation. Ferdie Wang wrote a doctoral dissertation developing new software technology for nonparametric regression. [24].

Roland Sanchez wrote a doctoral dissertation in which a nonparametric regression algorithm is developed for the short term forecasting of ozone levels. [9].

Nonparametric regression and hazard function estimation have been considered by Thompson as exploratory devices leading to the construction of explanatory models. The empirical smooths of Tukey are related to other nonparametric regression techniques. Thompson shows how hazard estimation from data sets from the Sixteenth Century to the present can be used to obtain insights and, frequently, models of the generating mechanism of these data sets. [21].

Resampling Strategies. Malcolm Taylor of BRL continues his joint work with Thompson on nonparametric density estimation motivated alternatives to the bootstrap. They have shown how the granularity pathologies associated with the bootstrap can easily be overcome with the fundamentally different (from the bootstrap) SIMDAT algorithm. [16].

Simulation Based Estimation. Detailed implementations of the SIMEST algorithm are presented by Atkinson, Brown and Thompson. [18]. An invited address to the Sesquicentennial Meeting of the ASA on the subject of SIMEST was given by Thompson.[20]. Ensor and Bridges have worked with Thompson in using SIMEST for the estimation of the parameters of a stochastic process characterizing the appearance and disappearance of hardware in the personal computer market. [7]. The parallelization of the SIMEST algorithm has been implemented by Stivers, Ensor and Thompson on a LEVCO transputer desktop system [23]. For the simulation based strategies considered,

the 16 transputer LEVCO works at roughly one-third the speed of a processor on a CRAY XMP, even though the latter machine costs roughly 200 times as much.

Modeling. Stivers and Ensor have worked with Thompson to examine his conjecture that "chaos" is generally an artifact caused by using deterministic models to model stochastic phenomena. They demonstrate how the chaos phenomenon disappears in two of the classical models of Lorenz when a tiny amount of noise is added. [23].

The possibility that closings of the gay bathhouses in the United States might have stopped the AIDS epidemic early on is considered. [17], [19]. Thompson and Go have considered what benefits might yet be available if public health intervention in the form of bathhouse closings was employed in the United States. It appears that as long as the fraction of HIV infectives in the high risk population is below 40% such closings would lower significantly the number of fatalities from AIDS. They have also considered the possibly negative effects (on the epidemic) of the use of AZT therapy. [22].

Personnel Supported.

Faculty: S.T. Chiu, Katherine B. Ensor, David W. Scott, James R. Thompson.

Graduate Students: Kerry Go, Thomas Kauffman, Martin Lawera, Roland Sanchez Martin Spears, David Stivers, Ferdie Wang.

Doctoral Degrees Awarded to Supported Students

Roland Sanchez (1990)

Ferdie Wang (1990)

Publications

(1) Chiu, S.T. (1988). "Weighted least squares estimators on the frequency domain for the parameters of a time series, *Ann. of Statist.*, **16,** 1315-1326.

(2) Chiu, S.T. (1989). "Detecting periodic components in a white time series," *JRSS Ser. B.*51 pp 249-259 (1989).

(3) Chiu, S.T. (1990). "Bandwidth selection for kernel estimation with correlated Noise," *Statistics and Probability Letters*. 8 pp. 347-354.

(4) Chiu, S.T. (1990). "On the asymptotic distributions of bandwidth estimates," *Ann of Statist.*, 18. pp 1696-1711.

(5) Chiu, S.T. (1990). "Peak-Insensitive parametric spectral estimation," *Stochastic Processes and their Applications*. pp 121-140.

(6) Ensor, Katherine B. (1990). "A recursive in order algorithm for least squares estimates of an autoregressive process," *J. of Statistical Computation and Simulation*, 37, pp 115-126.

(7) Ensor, Katherine B., Bridges, Eileen, and Thompson, James R. (1992). "Marketplace competition in the personal computer industry," accepted by *Decision Sciences*.

(8) Haerdle, Wolfgang and Scott, David W. "Smoothing in low and high dimensions by weighted averaging using rounded points." Rice Technical Report 88-16.

(9) Sanchez, Roland. (1990). *A Nonparametric Regression Algorithm for Time Series Forecasting Applied to the Predictive Analysis of Daily Maximum Urban Ozone Levels*. Doctoral dissertation, Rice University.

(10) Scott, David W. and Hall, M.R. (1989). "Interactive multivariate density estimation in the S Language." *Proceedings of the 20th Interface of Computer Science and Statistics*. American Statistical Association, pp. 241-245.

(11) Scott, David W. "Software for cross-validation of density estimates," Rice Technical Report 88-8-311.

(12) Scott, David W. and Wand, Matt P. "Feasibility of multivariate density estimates."*Biometrika*, 78:197-206.

(13) Scott, David W. (1990) "Statistical data analysis: How far will computer graphics take us?" *Proceedings of the Thirty-Fifth Army Design of Experiments Conference*, ARO, Research Triangle Park, August 1990.

(14) Scott, David W. "Some reflections on estimation and visualization of multivariate density functions," to appear, *Proceedings of the IMA, 1990.*

(15) Scott, David W. (1990). "Statistics in motion: where is it going?" *ASA 1989 Proceedings of Statistical Graphics Section,* 1990, pp. 17-22, ASA Alexandria, VA.

(16) Taylor, Malcolm S. and Thompson, James R., (1992). "A nonparametric density estimation based resampling algorithm." To appear in *Proceedings of the Bootstrap Conference*. LePage and Billard, eds.

(17) Thompson, James R. (1988). *Emprical Model Building*. John Wiley and Sons, 272 pages, ISBN 0-471-60105-5.

(18) Thompson, James R. , Atkinson, E. Neely and Brown, Barry W. (1989)."SIMEST and SIMDAT: differences and convergences," in *Proceedings of the Twentieth Interface Conference on Statistics and Computing Science*, E. Wegman, ed., Amsterdam: North Holland, pp. 581-586.

(19) Thompson, James R. (1989). "AIDS: the mismanagement of an epidemic," *Computers and Mathematics with Applications,* v. 18, pp. 965-972.

(20) Thompson, James R. (1989)."Microaxiom based simulation: a modern alternative to the closed form," in *Proceedings of the American Statistical Association: Sesquicentennial Invited Paper Sessions*, Ed. M. Gail and N.L. Johnson, pp. 259-271.

(21) Thompson, James R. and Tapia, Richard A. (1990). *Nonparametric Function Estimation, Modeling, and Simulation.*. Philadelphia: SIAM. 304 pages, ISBN 0-89871-261-0.

(22) Thompson, James R. and Go, Kerry, "AIDS: Modeling the mature epidemic in the American gay community," (1991). *Mathematical Population Dynamics,* Arino, Axelrod and Kimmel, eds., New York: Dekker, pp. 371-379.

(23) Thompson, James R., Stivers, David, and Ensor, Katherine B. (1991). "SIMEST: A technique for model aggregation with considerations of chaos," *Mathematical Population Dynamics,* Arino, Axelrod and Kimmel, eds., New York: Dekker, pp. 483-510.

(24) Wang, Ferdie. (1990). *A New Method for Robust Nonparametric Regression.* Doctoral dissertation, Rice University.