REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 10, 2005	Final Progress Report July 01, 2001 – Jan 31, 2005

4. TITLE AND SUBTITLE

Problems in Mathematical Statistics

5. FUNDING NUMBERS

DAAD19-01-1-0684;

6. AUTHOR(S)

Professor N. Rao Chaganty, Principal Investigator

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

Old Dominion University Research Foundation
P.O. Box 6369, Norfolk, VA 23508-0369

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

41879.13-MA

200 SUPPLEMENTARY NOTES

The views, 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.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

N/A

200 ABSTRACT (Maximum 200 words)

The primary goal of this project was to develop new statistical methods to meet the challenges of the evolving data analysis problems that the army encounters. These statistical methods also have numerous applications in biology, medicine and related sciences. Our research is focused on the following three important problems: (1) study mathematical details of the quasi-least squares method that we have developed for analyzing longitudinal and clustered data, (2) develop bivariate models for gene expression data to identify differentially expressed genes in microarrays, (3) study invariance properties of test statistics that occur in multivariate analysis of variance. For binary longitudinal data, we have studied the efficiency of generalized estimating equations with respect to a latent variable model and made some recommendations on how to choose the weight matrix. For continuous longitudinal data, we have studied theoretical properties of the quasi-least squares method. We have proved fairly general theorems establishing the asymptotic distributions of the quasi-least squares estimates. Using the asymptotic relative efficiency criterion we have shown that the quasi-least squares estimates are good competitors to the traditional maximum likelihood estimates obtained under the normality assumption for autoregressive time series regression models. In recent years, technology has made it possible to study gene expressions of thousands of genes simultaneously through the use of microarrays. We have developed novel statistical models to identify differentially expressed genes in microarray experiments. Next, for normal data we have obtained simplified versions of the Cochran's theorem for the independence and Wishartness of matrix quadratic forms for arbitrary covariance matrix. The results were used to characterize the class of covariance matrices such that the distributions of popular multivariate test statistics remain invariant except for a scale factor.

200 SUBJECT TERMS	200 NUMBER OF PAGES
quasi-least squares, microarrays, gene expressions, matrix quadratic forms.	200 PRICE CODE

200 SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	200 SECURITY CLASSIFICATION OF ABSTRACT	200 LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Problems in Mathematical Statistics

FINAL PROGRESS REPORT

March 10, 2005

U. S. Army Research Office

DAAD19-01-1-01684
41879-MA

Old Dominion University Research Foundation
P.O.Box 6369
Norfolk, Virginia 23508.

Approved For Public Release;
Distribution Unlimited.

Table of Contents

1 Statement of the problems studied.

The primary goal of this project was to develop new statistical methods and study their theoretical underpinnings to meet the challenges of evolving data analysis problems that the army encounters. These statistical methods also have applications in biology, medicine and related sciences. Our research is focused on three important problems: (1) study mathematical details of a new statistical method that we have developed for analyzing longitudinal and clustered data, (2) develop bivariate models for gene expression data to identify differentially expressed genes in microarrays, (3) study invariance properties of test statistics that occur in multivariate analysis of variance.

2 Summary of the most important results.

2.1 Quasi-least squares.

A major accomplishment of this project is the development of a new statistical method for analyzing longitudinal or repeated measurements data. Such data naturally occur when repeated observations are taken on individuals, or the data is taken on clusters or groups of subjects sharing similar characteristics. In a landmark paper, Liang and Zeger (1986, *Biometrika*, **73**, 13-22) introduced the generalized estimating equations (GEE) for analyzing longitudinal data. The GEE method has become so popular that the 1986 article of Liang and Zeger was included in Volume 3 of "Breakthrough in Statistics." But despite its popularity, the method has some significant problems, particularly in estimation of correlations between the repeated measurements. In this project we developed an improved method of estimating the correlations; the quasi-least squares method. Using the asymptotic relative efficiency criterion we have shown that the quasi-least squares estimates are good competitors to the maximum likelihood estimates obtained under the assumption of normality. Further, using simulations we have shown that the quasi-least squares estimates are robust and insensitive to the assumption of normality. They are undoubtedly better than the moment estimates.

2.2 Analysis of microarray data.

In recent years, technology has created a major revolution in biology and microbiology research. The revolution was made possible by the extensive use of the new inexpensive and high throughput chips, for example protein chips, mass spectrometry and microarrays. In particular, advances in microarray technology are enabling researchers to quantitatively analyze expression levels of thousands of genes simultaneously. During the early years of microarray studies, researchers relied mainly on traditional methods for analyzing gene expression data, including but not limited to, hierarchical clustering and t-tests. These methods, known to perform well for small data sets, have been only partially successful for large data. In this project we have developed new models to analyze gene expressions from microarrays. Our models, which account for the correlation between measured intensities of the control and cancerous tissues, are useful to calculate the posterior odds of gene expressions, and to select highly differentially expressed genes.

2.3 Matrix quadratic forms.

The popular statistical tests in multivariate analysis of variance are based on Cochran's theorem, which assumes that the samples are taken independently from normal populations. Much research has been done on the extensions of Cochran's theorem for matrix quadratic forms. However, these extensions still assume that observations within the samples are independent. In this research project we derived simple versions of the Cochran's theorem when the observations within each sample are correlated with covariance matrix Σ. In particular we have derived necessary and sufficient conditions such that common matrix quadratic forms are independent and distributed as Wishart in the cases where Σ is the Kronecker product of two nonnegative definite matrices, and an arbitrary nonnegative definite matrix. We have used the results to characterize the class of nonnegative definite matrices such that the matrix quadratic forms that occur in multivariate analysis of variance are independent and Wishart except for a scale factor.

3 List of all publications.

3.1 Papers published.

1. Analysis of growth curves with patterned correlation matrices using quasi-least squares. *Journal of Statistical Planning and Inference*, **117**, pp 123-139, 2003.

2. Wishartness and independence of matrix quadratic forms for Kronecker product covariance structures. (with A. K. Vaish). *Linear Algebra and its Applications*, **388**, pp 379-388, 2004.

3. A note on the estimation of autocorrelation in repeated measurements. (with G. Shi). *Communications in Statistics: Theory and Methods*, **33**, pp 1157-1170, 2004.

4. Application of quasi-least squares to analyze replicated autoregressive time series regression models. (with G. Shi). *Journal of Applied Statistics*, **31**, pp 1147-1156, 2004.

5. Bivariate models for identifying differentially expressed genes in microarray experiments. (with D. Mav). *Journal of Statistical Theory and Applications*, **3**, pp 111-124, 2004.

6. Efficiency of generalized estimating equations for binary responses. (with H. Joe). *Journal of the Royal Statistical Society: Series B*, **66**, pp 851-860, 2004.

3.2 Papers tentatively accepted for publication.

1. Nonnegative definite solutions to matrix equations with applications to multivariate test statistics. (with A. K. Vaish). Submitted to *Linear Algebra and Its Applications.*

3.3 Technical reports.

1. On efficient estimation of the regression parameter for correlated data in generalized linear models. (with J. Shults).

4 Degrees awarded.

1. Genming Shi, Phd 2003. Thesis title: "Estimation of parameters in replicated time series regression models."

5 Presentations and invited talks.

1. Wishartness and independence of quadratic forms in correlated singular normal vectors. 10th International workshop in matrices and statistics, Voorburg, The Netherlands, August 2-3, 2001.

2. Statistical analysis of some multivariate repeated measurement models. Department of Biostatistics, Penn State University, Hershey, June 2001.

3. Computational Aspects of Medical and Pharmaceutical Statistics, Discussant of Invited Papers, JSM-ASA, August 14, 2002, NY.

4. Analysis of correlated data using estimating equations, Dept of Statistics and Actuarial Science, Nov 14, 2002, University of Waterloo, Canada.

5. Statistical methods in bioinformatics, May 20, 2002, Old Dominion University.

6. Statistical analysis of gene expression in microarray experiments, Colloquium, October 23, 2002, Old Dominion University.

7. A note on the estimation of autocorrelation in repeated measurements, Hawaii International Conference on Statistics and Related Fields. June 5-8, 2003, Honolulu, Hawaii.

8. Nonnegative definite solutions to matrix equations with applications to multivariate statistics, 12th International Workshop in Matrices and Statistics, August 5-8, 2003, Dortmund, Germany.

9. A mixture model for the analysis of correlated binomial data, ENAR 2003, March 30-April 2, 2003, Tampa, FL.

10. Bivariate models for identifying differentially expressed genes in microarray experiments, Graybill Conference, June 2003, Fort Collins, CO.

11. A mixture model for the analysis of correlated binomial data, May 2003, Virginia Academy of Sciences, Charlottesville, VA.

12. Application of quasi-least squares to analyze replicated autoregressive time series model, 67th Annual Meeting of the IMS, University of Barcelona, Spain, July 26 - 31, 2004.

13. Generalized estimating equations: What's wrong with them? Virginia Academy of Sciences, May 2004, Richmond, VA.

14. Generalized estimation equations: A critical review. Colloquium, September 17, 2004, Virginia Commonwealth University, Richmond, VA.

15. Asymptotic behavior of statistical methods for the analysis of correlated Poisson outcomes, International conference on Future of Statistical Theory and Practice, December 2004, Hyderabad, India.

6 Professional activities.

1. Chaired a Session at the Virginia Academy of Sciences Meetings, Hampton, VA, May 2002.

2. Chaired a Session at the IISA meeting, DeKalb, IL, June 2002.

3. Chaired a Session at the Virginia Academy of Science Meetings, Charlottesville, VA, May 2003.

4. Organized and Chaired a Session at International conference on Future of Statistical Theory and Practice, December 2004, Hyderabad, India.

7 Participating scientific personnel.

Narasinga Rao Chaganty, Principal Investigator.

Genming Shi, Graduate Student.

Deepak Mav, Graduate Student.