METHODS IN PROBABILITY AND
STATISTICAL INFERENCE

Final Report
for Period 15 June 1975 - 30 June 1979

David L. Wallace, Principal Investigator (1977-1979)
Michael D. Perlman, Principal Investigator (1975-1977)

Department of Statistics
University of Chicago
Chicago, Illinois 60637

MASTER

June 1980

Prepared for

THE DEPARTMENT OF ENERGY
OFFICE OF BASIC ENERGY SCIENCE
DIVISION OF ENGINEERING, MATHEMATICAL AND GEOSCIENCES
UNDER CONTRACT NO. EY-76-S-02-2751.*000

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed or represents that its use would not infringe privately owned rights.

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ABSTRACT

I. Statistical Computing; Approximation Techniques in Statistics.

David L. Wallace has extended the development of the SNAP statistical system to provide flexible interactive data analysis and graphical display. He has developed various approximations and explored their applications to the characteristic roots of Wishart matrices, to Mandel-Golub mixed additive multiplicative models, to inhomogeneous k-statistics, to the Behrens-Fisher problem and to simplified probability plotting.

II. Numerical Computation of First Passage Distributions.

Michael J. Wichura has implemented numerical procedures for computing first passage distributions to a moving boundary for various diffusion processes, including the Wiener process and Ornstein-Uhlenbeck process. Procedures for dealing with various kinds of singularities in the first passage distribution have been developed. Similar work has been carried out for extremal processes.

III. Probabilities of Large Deviations.

R. R. Bahadur and Sandy Zabell have studied the application of certain ideas from statistical mechanics to the estimation of large deviation probabilities. They have developed a unified theory for large deviations of the sample mean of independent and identically distributed random vectors taking values in a vector space. The theory is sufficiently general to include the case when the summands are Brownian motions or bridges, and the case when the sample mean is the empirical distribution function.

IV. Combining Independent Tests of Significance.

Michael D. Perlman and co-author James A. Koziol have carried out a comparative study of classical nonparametric "omnibus" and Bayes procedures for combining several independent noncentral chi-square or F test statistics. Perlman and doctoral student John Marden obtained general necessary and sufficient conditions for the admissibility of combination procedures. R. R. Bahadur has investigated combination procedures from the point of view of second order asymptotic efficiency.

V. Small-Sample Efficiencies of Tests and Estimates.

R. R. Bahadur, James A. Koziol, and David L. Wallace have studied the exact performance of likelihood ratio and chi-square

tests of a simple trinomial hypothesis. They have demonstrated that some recent asymptotic comparisons are seriously misleading for practice and suggest improved forms for representing and comparing behavior of these tests.

VI. Improved Procedures for Simultaneous Estimation and Testing of Many Correlations.

Michael D. Perlman and doctoral student S. P. Lin have developed a new technique for the efficient simultaneous estimation of all correlation coefficients in a multivariate population. Their method adapts the James-Stein "shrinking" estimator (for location parameters) to the estimation of correlations. Perlman and doctoral student John Marden have investigated testing for zero correlation under various side conditions, and for testing hypotheses about a multivariate mean under covariates. Perlman and co-author K. Larntz have developed a simple test for the equality of several correlation matrices.

VII. Statistical Computing and Improved Regression Methods.

Ronald Thisted has studied the evaluation of statistical software, development of statistical algorithms, and possible improvements to least-squares regression in the presense of multicollinearity. In the first area he has developed criteria for software evaluation, has showed how effective software utilization can be incorporated in the training of data analysts, and has developed guidelines for effective documentation of statistical algorithms. In the second area, Thisted has developed new algorithms for isotonic modal regression estimators and for computing the distribution of a family of two-sample rank tests, and he has developed a computer program for data analysis using ridge regression methods. Finally, he has conducted a thorough investigation of the behavior of ridge regression estimators, has shown conditions under which they perform well (and others where they do badly), and has examined theoretical properties of ridge and similar estimators.

VIII. Comparison of Several Populations.

James A. Koziol has developed techniques for making multiple comparisons among several populations differing by a scale parameter, such as exponential distributions with varying mean-lifetime. He has also developed a multiple comparison technique for comparing the means of several populations when the samples involved are subject to unequal patterns of censorship; this technique has application to the analysis of incomplete survival data. He has prepared an additional paper pertaining to the sequential analysis of incomplete survival data. Further, he has tabulated the null distribution of a certain nonparametric test for bivariate symmetry.

IX. Unbiasedness in Multivariate Statistics.

Michael D. Perlman has developed new methods in order to resolve several open questions concerning the unbiasedness of tests in multivariate analysis. Tests considered include Bartlett's modification of the likelihood ratio test for equality of the covariance matrices of several multivariate normal populations, the class of all admissible invariant tests for the multivariate analysis of variance, and the analogous class of tests for testing independence of two sets of variates.

CONTENTS

	<u>Page</u>
A. INTRODUCTION	1
B. FINAL RESEARCH REPORT: 15 June 1975 - 30 July 1979	2
I. Statistical Computing; Approximation Techniques in Statistics.	2
Investigator: David L. Wallace	
II. First and last passage distributions to moving boundaries by stochastic processes.	5
Investigator: Michael J. Wichura	
III. Probabilities of Large Deviations.	7
Investigators: R. R. Bahadur and Sandy Zabell	
IV. Combining Independent Tests of Significance.	9
Investigators: Michael D. Perlman, James Koziol, and R. R. Bahadur	
V. Small-sample Efficiencies of Tests and Estimates.	11
Investigators: R. R. Bahadur, James A. Koziol, and David L. Wallace	
VI. Improved Procedures for Simultaneous Estimates and Tests of Many Correlations.	13
Investigator: Michael D. Perlman	
VII. Statistical Computing and Improved Regression Methods.	16
Investigator: Ronald Thisted	
VIII. Comparison of several populations.	19
Investigator: James A. Koziol	
XI. Unbiasedness in Multivariate Statistics.	22
Investigator: Michael D. Perlman	

C. CONSULTATION AND COOPERATIVE RESEARCH WITH OTHER SCIENTISTS AT THE UNIVERSITY OF CHICAGO AND ARGONNE NATIONAL LABORATORIES	23
D. PUBLICATIONS ISSUED: 15 June 1975 - 30 July 1979	28

A. INTRODUCTION

This report describes the research activities of the Department of Statistics, University of Chicago, carried out or in progress under the auspices of DOE Contract No. EY-76-S-02-2751.*000, during the period 15 June 1975 to 30 July 1979. Nine research projects are described in Part B of this report. Part C contains a description of the statistical consultation activities of the Department which are of interest to DOE, in particular the scientific interactions between the Department and the scientists at Argonne National Laboratories. A list of publications issued during the term of the contract appears in Part D.

B. FINAL RESEARCH REPORT: 15 June 1975 - 30 July 1979

I. Statistical Computing; Approximation Techniques in Statistics

Investigator: David L. Wallace

(a) Statistical Computing.

Developments in statistical computing were aimed at encouraging flexible data-analysis, especially through interactive analysis and graphical output. Work centered on the SNAP statistical computing system that Wallace had developed over the past decade. Throughout the period of report, the SNAP system continued in heavy use for statistical instruction and for research in theoretical statistics and in applied fields where more flexibility and control over choice of analysis is needed than SPSS and other packages provide. The ability for easy construction and testing of tailored methods of analysis is an important feature.

One major development was the incorporation of interactive graphics based on the TEKTRONIX 4010 series of graphics terminals within the IBM, and later the DEC, versions of SNAP. For practical data analysis, even a rich set of graphing subroutines, such as the TEKTRONIX PLOT-10 package, must be combined with an interactive control program and adequate data-analytic capabilities. Under the TSO time sharing option on the IBM 370/168 computer, the open-ended structure of the SNAP system provided a convenient and direct implementation for the graphing operations. Additionally, any user can incorporate special graphical and analytic operations into a private version of the SNAP system.

The 1978 installation of a DEC20 computer for academic computing led to the conversion of the SNAP system and its enhancement for more effective interactive computing in the friendly DEC20 environment. Analytic capabilities added during the project period include maximum likelihood estimation of missing values in multivariate data, a variety of quantal response techniques, harmonic analysis commands, improved analysis of factorial designs via generalized Yates' algorithm. Other improvements include commands to facilitate the easy transfer of data between SNAP and other programs and systems, and (only on the DEC20) facilities to allow the interpolation of FORTRAN computations into SNAP processing.

(b) Approximation Techniques in Statistics.

A variety of problems were explored with many partial results not yet ready for publication.

(i) Inhomogeneous k-statistics.

Fisher named the "k-statistics" and provided an extensive technology for working out their sampling cumulants. Stimulated originally by a problem in

stellar astronomy, Wallace developed the sampling properties of k-statistics under inhomogeneous sampling, where different parts of the sample come from different populations, with the goal of providing usable tools for approximating behavior of standard data-analytic procedures that are applied to inhomogeneous data. Results obtained for the first two orders of cumulants are vastly longer than, the homogeneous results, but with pleasing structure.

(ii) Simplified probability plotting.

Probability plotting is a common and useful exploratory and diagnostic technique whose value and general utility have often been limited by an inappropriate insistence on a precise choice of plotting positions, by the need to estimate parameters, and a necessity for specially designed graph paper or else extensive tables or non-trivial computer programs. Gamma (or chi-square) plotting illustrates the approach based on approximate techniques.

The Wilson-Hilferty cube root transformation of chi-squares provides an approximate normalization that is convenient and precise enough for graphing even down to one degree of freedom. If Y is purportedly distributed like $c\chi^2_\nu$, then

$$\sqrt[3]{Y} \approx \sqrt[3]{c\nu} \left(1 - \frac{2}{9\nu}\right) + \sqrt[3]{c\nu} \sqrt{\frac{2}{9\nu}} Z$$

where Z is unit Gaussian. Making an ordinary full normal plot of the cube-roots of a sample should produce nearly a straight line with intercept and slope simply related to c and ν . Linearity of the plot is a check on Y being a multiple of chi-square, and graphical estimates of c and ν are easily obtained. Often ν is known, but c is unknown, and in this case the plot should correspond to a line with prescribed intercept on the Z -scale.

Sampling variation prevents even truly normal samples from yielding straight lines. For truly chi-square samples, the slight curvature introduced because of the inexactness of the cube root normalization will be small relative to sampling fluctuation except for very small degrees of freedom or very large samples or in the (often uninteresting) lower tail of the sample.

(iii) Approximating distributions of characteristic roots.

A class of approximations to moments and percent points of the first root of a Wishart matrix were developed. Desiderata include convenient analytic form for representation in computers and for interpolation to untabled values. The form adopted is a mixed additive and multiplicative form, of the type fitted in Mandel-Gollob models, whose theory in turn requires the distribution of the roots of the Wishart matrix. The fitted function is a symmetrical function of two integer arguments, whose margins for either argument one are elementary. This fact is used to choose the additive term, rather than using a least-squares fit. The choice of multiplicative terms were sometimes based on empirical fit, sometimes using second margins, but all choices stress analytic convenience.

A preliminary transform of the root greatly improves the precision of the fit. The square root transform is nearly best, and greatly aids the interpretability of the result, in application to simultaneous confidence intervals, for example.

- (iv) Approximations to compromise solutions to the Behrens-Fisher problem.

For practical implementation, all solutions to the Behrens-Fisher problem other than the "Student" solutions, require approximations. For the Welch and Behrens solutions, the approximations are classical, though for the latter, improvements beyond Cochran's are desirable, and for the former, sensible extrapolation to small samples invalidates the favored solutions. Wallace has proposed a family of Bayesian compromise solutions based on what he calls a Cornfield limiting prior. Explorations of behavior and practical approximations for using the procedure are under construction. The critical review paper ([5]) on the Behrens-Fisher problem presents a part of this work.

References

- [1] Wallace, David L. (1979). SNAP Users Manual, October 1979, 116 pages, University of Chicago Computation Center On-line Documentation.
- [2] Wallace, D. (1979). Plotting on the TEKTRONIX in DEC-SNAP, SNAP Memorandum Number 10, July 79(Rev).
- [3] Wallace, D. and Wichura, M. (1979). PSAVE and PEND: Routines to save data in SNAP's SAVE format, SNAP Memorandum Number 11, September 1979.
- [4] Wallace, David L. (1980). Interpolating FORTRAN computations within SNAP, SNAP Memorandum Number 12, July 1980.
- [5] Wallace, David L. (1980). The Behrens-Fisher and Fieller-Creasy problems. in Fienberg, S.E. and D.V. Hinkley (Editors), R. A. Fisher: An Appreciation, Lecture Notes in Statistics, 1, pp. 119-147. New York: Springer-Verlag.
- [6] Wallace, David L. (1980). Analytic approximations to distributions of largest eigenvalues. In preparation.
- [7] Wallace, David L. (1980). Gamma probability plotting can be simple. In preparation.
- [8] Wallace, David L. (1980). Sampling cumulants of multivariate generalized k-statistics for inhomogeneous sampling. In preparation.

II. First and last passage distributions to moving boundaries by stochastic processes.

Investigator: Michael J. Wichura

Consider a system, or particle, evolving randomly in time. A problem of widespread interest in the natural sciences, in reliability theory, in statistics, and in still other disciplines, is to determine the first passage distribution of the system to various moving boundaries. To cite but one example, a sequential procedure for testing whether the system is stable, i.e., running without drift, or unstable, i.e., drifting out of control, typically involves moving boundaries chosen so that the probability of passage to the boundaries is small if the process is stable, but is ever larger the more unstable is the system. A common feature of these first passage problems is that, except in rare cases, no closed form solution is available, and numerical calculations are the only means of obtaining an exact solution.

(a) Numerical computation of first passage distributions for diffusions.

Wichura [1] has implemented a computational procedure for the rapid and efficient determination of first passage distributions to moving boundaries by diffusion processes, such as the Wiener process and the Ornstein-Uhlenbeck process, for which the transition densities of the process can be evaluated without undue expense. Absorption may be "one-sided", involving a single boundary, or "two-sided", involving a pair of boundaries. First passage densities are obtained by solving Fortet's Volterra integral equation of the first kind, using techniques of Anderssen-deHoog-Weiss and H. Brunner. First passage distributions are obtained by quadrature.

The program utilizes flexible input and output procedures which allow the user considerable freedom in setting up the problem and outputting the results of the computation. The user may select one of several "in house" families of parameterized moving boundaries, or provide his own. Similarly, the user may select one of several "in house" parameterized diffusions, or provide his own. The output routines permit the user to determine the amount of information printed and to store the results for future analysis. A computer graphics routine provides immediate visual displays of first passage densities and distribution functions, and boundaries. Facilities for calculation of relevant summary statistics, such as first passage moments, modes, and percentiles are incorporated.

The computational algorithm is predicated, roughly speaking, on a "smooth" boundary and a "smooth" first passage density. However, as experience has shown, it is not uncommon for the first passage density to have a "near" singularity, e.g. a pronounced narrow mode, or to have a genuine singularity, e.g. an explosion. Such phenomena typically cause the procedure to labor, in the first case, and to break down, in the second case. Various methods have been designed to cope with such difficulties. One is to use transformations of time, which preserve the diffusion property (albeit altering the infinitesimal velocity and variance) while, when properly chosen, mollify the singu-

larity. Another is the use of adaptively chosen variable panel widths for the piecewise polynomial bases used. Both methods have proven useful. The second method, being automatic, is more convenient, although somewhat more costly. The two methods may be used in conjunction.

(b) Passage times for the Wiener process.

There is a class of first passage problems for the Wiener process W which present singularities so severe that computational procedures based on Fortet's equation, even with tricks of transformation, and the like, do not seem to be able to cope. These problems arise in calculating boundary crossing probabilities of the form

$$P[W(t) > g(t) \text{ for some } t > s];$$

such probabilities could arise, e.g., in connection with the chance that the random system W should ever exceed some moderately growing safety threshold. The standard time inversion for the Wiener process converts this last crossing problem to a first passage problem.

Wichura has developed a theoretical framework within which the numerical solution of this first passage problem is feasible. The key ingredient in the method is an integral equation radically different from Fortet's. Asymptotic analysis of this equation, a Volterra equation of the second kind, has given a suitable basis for expanding the solution. Standard methods for solving equations of the second kind may be employed, subject only to devising quadratures to handle the extremely singular kernel. Work on this project is continuing, and will appear in [2].

(c) Passage times for extremal processes.

Wichura has also developed techniques for calculating boundary crossing probabilities for extremal processes, which stand in the same relation to the Wiener process as do extreme values to sums. The approach here parallels that described in the preceding paragraph, and has progressed to the point of providing simple approximations, e.g. double exponential, to the calculated distributions. This work is continuing, and will appear in [3].

References

- [1] Wichura, M.J. (1980). A computer program for the numerical evaluation of first passage distributions. In preparation.
- [2] Wichura, M.J. (1980). Boundary crossing probabilities for the Wiener process. In preparation.
- [3] Wichura, M.J. (1980). Boundary crossing probabilities for extremal processes. In preparation.

III. Probabilities of Large Deviations.

Investigators: R. R. Bahadur and Sandy Zabell

A central concern in statistical theory has always been the study of the efficiency of proposed methods of estimation and hypothesis testing. During the last twenty years, various statisticians (Bahadur, Chernoff, Hoeffding, and others) have proposed measures of efficiency which depend on accurate mathematical determination of extremely small probabilities. The mathematical problem involved here is in the area known as "large deviations", and the requirements of statistical theories have stimulated diverse investigations in the large deviations area. The most important single estimator is, perhaps, the sample mean. Bahadur and Zabell have made a very general study of large deviations of this particular statistic, along the following lines.

Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random vectors taking values in a space V , let Z_n be the sample mean of the first n X 's, and for $J \subset V$, let

$$a(J;n) = \log P[Z_n \in J] / n \quad (1)$$

for $n = 1, 2, \dots$. The set J is said to have entropy $S(J)$ if

$$\lim a(J;n) = S(J) \quad (2)$$

as n tends to infinity.

Questions concerning the existence and evaluation of $S(J)$ were initiated by Cramer (1938) and Chernoff (1952) in the case when the X 's are real valued and J is an infinite interval. The initial work of Cramer and Chernoff has been continued and extended in several directions. A seemingly different (but in fact formally equivalent) area of investigation, concerning large deviations of the empirical probability distribution on V was initiated by Sanov (1957).

A fresh and powerful attack on the validity of (2) was made by Lanford (1971) in the case when the X 's are real valued and bounded. Landford's approach is based on certain ideas from statistical mechanics concerning relations between micro- and macro-canonical ensembles. Bahadur and Zabell [1] have shown that Lanford's approach is successful even in the general case. Some elements of the theory are outlined below.

Suppose that V is a locally convex real vector space and that some very weak additional conditions are satisfied. Then $S(J)$ exists for every finite union of open convex sets. For each v in V , define the point entropy $s(v)$ in terms of the set entropy function S by

$$s(v) = \inf\{S(J): J \text{ is open, convex, and contains } v\}. \quad (3)$$

Then

$$S(J) = \sup\{s(v): v \text{ in } J\} \quad (4)$$

for every finite union of open convex sets.

The following is one method for finding the point entropy function s explicitly. Let V' be the set of all continuous linear functionals f on V . Let C be the cumulant generating function of the common distribution of the summands, i.e.

$$C(f) = \log E(\exp[f(X)])$$

for f in V' . Then

$$s(v) = \inf\{f(v) - C(f) : f \text{ in } V'\}. \quad (5)$$

In case V is the space of continuous functions v on $[0,1]$, and the summands are standard Brownian motion, $s(v) = -\int (dv/dt)^2 dt$ if v is absolutely continuous and the integral is finite, and $s(v) = -\infty$ otherwise. In case V is the space of finite signed measures on a Polish space Z , and the X 's are unit masses at independent identically distributed points Y_i in Z , then $s(v) = -\infty$ if v is not a probability measure, while $s(v) = -K(v,p)$ otherwise, where p is the actual probability distribution of Y and $K(v,p)$ is the average Kullback-Leibler information for testing p against v when the sample point is Y_1 .

Reference

- [1] Bahadur, R.R. and Zabell, S. (1979). Large deviations of the sample mean in general vector spaces. Annals of Probability. 7 pp. 587-621.

IV. Combining Independent Tests of Significance.

Investigators: Michael D. Perlman, James Koziol, and R. R. Bahadur.

If investigators at several, say n , different laboratories working independently have conducted experiments designed to test the same hypothesis, a statistical procedure is needed which will combine these n independent tests and yield an overall significance level concerning the validity of the common hypothesis. It is often the case that the test statistics to be combined follow chi-square or F-distributions, noncentral under the alternatives. In such situations combination procedures may be found which are more sensitive than "omnibus" nonparametric procedures such as Fisher's, which is based on the product of the attained significance levels.

Perlman and Koziol have carried out an intensive investigation of the problem of combining independent chi-square or F tests. Their major theoretical contribution has been the construction of several families of "conjugate" prior distributions for the noncentrality parameters, these families being wide enough to approximate any unimodal prior density fairly well, and yet easily inverted to obtain useable Bayes tests. They have also performed an extensive numerical study the power of these Bayes procedures and their various nonparametric competitors. They found that Fisher's procedure is approximately minimax among the procedures considered and furthermore closely approximates in power one or more Bayes procedures. When prior information is available about the magnitudes of the noncentrality parameters expected if the null hypothesis is false, one of the Perlman-Koziol Bayes combination procedures is preferable. The complete results of these investigations appear in [1] and [2].

In a subsequent study Perlman and J. Marden undertook a decision-theoretic analysis of these problems. For the parametric models mentioned above, they obtained necessary and sufficient conditions for admissibility of combination procedures. Whereas Fisher's procedure satisfies these conditions in each problem considered, other popular combination methods (such as those based on the sums of the inverse normal or inverse logistic transforms of the observed significance levels) usually do not, and so are inadmissible. This work appears in [3].

The preceding two paragraphs concern finite sample sizes. It has been known for some time that if large samples are considered, a variety of nonparametric combination procedures, including Fisher's, the inverse normal, and the inverse logistic mentioned above, are asymptotically optimal under the criterion of Bahadur efficiency. Bahadur has undertaken a careful study of these results and has found that when second-order terms are considered, Fisher's method is the unique optimal procedure under an asymptotic minimax criterion. This result nicely complements the numerically-based results of Koziol and Perlman mentioned earlier. This work appears in [4].

References

- [1] Perlman, M.D., and Koziol, J. (1978). Combining independent non-central chi-square tests. Journal of the American Statistical Association 73 pp. 753-763.
- [2] Perlman, M.D., Koziol, J., and Rasmussen (1980). Combining independent non-central F-tests. In preparation.
- [3] Perlman, M.D., and Marden, J. (1980). Combining independent chi-square and F tests: admissibility and complete class theorems. Technical Report, Department of Mathematics, Urbana.
- [4] Bahadur, R.R. (1980). On combining independent tests of significance. In preparation.

V. Small-sample Efficiencies of Tests and Estimates.

Investigators: R. R. Bahadur, James A. Koziol, and David L. Wallace

The likelihood-ratio test and the Pearson chi-square test are used widely in statistical practice. Classical asymptotic comparisons assert the asymptotic equivalence of the two tests in many situations. A more refined asymptotic analysis based on large deviations asserts, in some important situations where the analysis has been completed, that against almost every alternative the chi-square test is inferior asymptotically. Bahadur, Koziol, and Wallace undertook an empirical comparison of the exact performances of these two tests of a simple trinomial hypothesis for a sequence of small-to-large samples. The objective was in part to facilitate choice between these tests, but more broadly to develop methods for bridging the gap between exact, small-sample computations and theoretical asymptotic behavior. The study demonstrates that some asymptotic comparisons provide misleading guides for practice and understanding; some alternative comparisons give a better basis for choice.

Consider a fixed null hypothesis and study the behavior of the tests against a temporarily fixed alternative hypothesis. For any sample size, n , the median of the distribution of the exact significance level achieved by each test is an important quantity; it is, equivalently, the type I error level required to yield power exactly one half. The negative logarithm, y_n , of the median significance level is the central object of study. A comparison of the rates at which y_n grows is the primary comparison of the tests; it is easily converted into a comparison of equivalent sample sizes required for comparable performance.

Asymptotic theory asserts that, $\lim y_n/n = c^e$ where c^e is the "exact half-slope" of the test statistic for the particular null and alternative hypothesis. Inasmuch as the exact half-slope for chi-square never exceeds that for the likelihood ratio statistic, the comparison of the limiting rate favors the latter test statistic. The exact comparisons over a wide range of sample sizes and alternative hypotheses that typically includes all situations where the significance level ranges from about .1 down to .00001 (the "range of interest") is in conflict with this asymptotic conclusion. Over this range of interest, the two tests are competitive, with nearly equal regions of alternatives in which each is superior to the other. A more careful data analysis clarifies the conflicts, and suggests needed refinements to the asymptotic theory for practical applicability.

For the likelihood ratio test, the behavior of the median y_n is well represented for all sample sizes by the leading term and an additive adjustment:

$$y_{n,LR} = c_{LR}^e n + a$$

The additive term varies somewhat with hypothesis, but is more close to .5. Deviations from this expression are equivalent, typically, to a change of less than one or two in the sample size. The behavior for the chi-square test is

more complex. Over the wide "range of interest", the median y_n for the chi-square test is also represented moderately well by an expression of the same linear form, except that the coefficient of n is NOT the exact half-slope for the chi-square statistic as specified by the leading asymptotic theory, but a quantity that is always larger. The asymptotic theory is not contradicted by the empirical study, for as the sample size gets very large corresponding to significance levels well below 10^{-5} , the y_n for chi-square starts to bend toward the required limit. The comparison of the "approximate" slopes of the fitted lines give a better basis for judging merits than do the exact slopes. Some other theoretical considerations suggest an analytic form for the "approximate" slope for chi-square.

Asymptotic theory suggests that the next term ought to be of order $\log n$, though the form and the coefficients cannot now be determined. Empirical modelling of the exact behavior leads to deeper understanding of behavior and theory. For the chi-square test, the median y_n is well represented by:

$$c_{CS}^e n + b \log(1+dn) + a$$

where all but the exact slope c^e are determined by empirical fit, and vary with the hypothesis. The fit for the likelihood ratio shows no need for a logarithmic term. The observed linear fit with an "approximate" slope is explained by the fact that in the range of interest, the log term is well approximated by its linear Taylor series, so the effective slope is $c^e + bd$, not c^e . The logarithmic term becomes relatively unimportant only for sample sizes and hypotheses that correspond to extreme significance levels.

In summarizing the findings, several points of substantive and methodological interest have been omitted. The discreteness of the underlying distributions causes what looks very much like a random noise superimposed on regular behavior. Its effects are usefully described by the root mean square amplitude of a pseudo random noise process, and typically amount to a jitter of about 1 extra observation. Secondly, for designs of studies, the spread of the distribution is important along with a center value. The semi-interquartile range is almost, subject to the jitter, directly proportional to square root of sample size, with a coefficient that depends only slowly on particular hypotheses and tests.

References

- [1] Bahadur, R.R., Koziol, J. and Wallace, D.L. (1980). Comparing test efficiencies--leading asymptotic terms can be misleading. In preparation.

VI. Improved Procedures for Simultaneous Estimates and Tests of Many Correlations.

Investigator: Michael D. Perlman

The problem of simultaneously estimating all correlations among the components of a multivariate population is extremely important, yet difficult if the dimension is not small and accurate estimates are required. The reason for this is quite simple: if the multivariate observations are vectors of dimension p , then there are $p(p-1)/2 \approx p^2/2$ correlations to be estimated, as opposed to just p means. Thus much larger sample sizes are required to obtain accurate estimates of the population correlation matrix as compared to estimates of the population mean vector. For small to moderate sample sizes, standard maximum likelihood estimates of correlation coefficients tend to be inaccurate and often too large in magnitude.

Perlman and doctoral student S. P. Lin have developed an improved technique for obtaining simultaneous estimates of correlation coefficients. Their idea is to first apply the Fisher z -transformation to each of the $p(p-1)/2$ correlation coefficients, then to apply a James-Stein Efron-Morris type estimation procedure, and finally to transform back via the inverse Fisher transformation. Extensive Monte-Carlo studies have shown this technique provides substantial improvement over the maximum likelihood estimator for various loss functions, for both moderate and small sample sizes.

The Perlman/Lin correlation technique has led to an improved method for the related problem of estimating the entire covariance matrix. Monte-Carlo studies have been carried out comparing maximum likelihood estimation, the Perlman/Lin correlation matrix method, and the characteristic roots method of Charles Stein. It was found that substantial improvement over the maximum likelihood estimator can be obtained by the characteristic roots method when the population characteristic roots are nearly equal, and by the correlation matrix method when magnitudes of the population correlations can be approximated a priori, for both moderate and small sample sizes. The Perlman/Lin correlation technique appears in detail in [1] and [2].

The need for simple yet accurate procedures to test the equality of two or more population correlation matrices has long been recognized. This problem arises, e.g., in biological applications and factor analysis. However due to the large number of correlations present in p -variate populations, and due to the computational complexities arising in the calculation of maximum likelihood estimates and the likelihood ratio test, up to now there have been no satisfactory procedures for small and moderate sample sizes. In 1967 Kullback proposed a simple test statistic and claimed an asymptotic chi-square distribution under the null hypothesis, but it was later shown that this distribution theory was incorrect, and that in fact the asymptotic distribution of Kullback's statistic depends on the unknown parameters. In 1970 Jennrich proposed a reasonable test statistic whose asymptotic null distribution is indeed chi-square. The accuracy of this approximation depends however on the approximate multivariate normality of the sample correlation matrix, and therefore requires

very large sample sizes. Furthermore, calculation of Jennrich's statistic requires the inversion of two $p \times p$ matrices.

Perlman and K. Larntz have developed a simple statistic for testing equality of several correlation matrices which requires no matrix inversion, may be computed by means of a hand calculator, and whose asymptotic distribution is approached much more rapidly than that of Jennrich's statistic. In the two sample case, the Perlman/Larntz statistic is based simply on the maximum absolute difference of the Fisher z -transforms of corresponding sample correlation coefficients. Monte-Carlo studies have been carried out comparing the performance of the Perlman/Larntz statistic with that of Jennrich's (and two modifications) in the null and non-null cases. As mentioned earlier, for small sample sizes the Perlman/Larntz statistic is clearly preferable to Jennrich's. For large sample sizes the Jennrich statistic is more powerful than the Perlman/Larntz statistic for some alternatives, while the situation is reversed for other alternatives. This work is presented in [3].

The problem of testing for zero correlation, i.e. independence, among normally distributed variates is not straightforward when observations are missing on some variates, and/or extra observations are available on others. This is just one of the problems addressed by Perlman's doctoral student J. Marden, who has carried out an extensive study of testing for zero correlation under a variety of side conditions. This work is part of Marden's thesis and appears in [4,5]. On a related matter, Perlman and Marden have completed their joint paper "Invariant tests for means with covariates" [6]. Here they obtained the minimal complete class of invariant tests for testing a hypothesis about a multivariate normal mean vector when covariates are present, and also made numerical comparisons among the tests which have been used for this problem.

References

- [1] Lin, S. P. (1978). An improved procedure for the estimation of a correlation matrix. Technical Report #69, Department of Statistics, University of Chicago.
- [2] Lin, S. P. and Perlman, M.D. (1980). A Monte Carlo comparison of four methods for estimating a covariance matrix. Technical Report, Department of Mathematical Sciences, Memphis State University. Submitted to Journal of the American Statistical Association.
- [3] Larntz, K. and Perlman, M.D. (1980). A simple procedure for testing equality of correlation matrices. In preparation.
- [4] Marden, J. (1978). Complete class theorems for invariant tests in multivariate analysis. Technical report #52, Department of Statistics, University of Chicago.
- [5] Marden J. (1980). Invariant tests on covariance matrices. Technical Report, Department of Mathematics, University of Illinois at Urbana. Submitted to Annals of Statistics.
- [6] Marden J., and Perlman, M.D. (1980). Invariant tests for means with covariates. Annals of Statistics 8 pp. 25-63.

VII. Statistical Computing and Improved Regression Methods.

Investigator: Ronald Thisted.

(a) Evaluation of statistical software.

Two articles have reported on the development of criteria for evaluating statistical software and of methods for helping scientists and teachers to use that software efficiently. One report [1] examined several statistical computer packages with emphasis on user documentation and control language. It is shown there how to supply guidelines to users to assist them in selecting packaged software best suited to their individual needs in solving problems. The other report [2] focused on training consumers to utilize packaged software and other computational resources effectively. Explicit guidelines for incorporating statistical software packages in the training of data analysts are given there as well.

A third report [3] examines the algorithms and algorithmic documentation for statistical procedures used in the SPSS statistical package. Aspects of documentation which make user evaluation more fruitful are discussed at length. Many of the recommendations in the report were adopted by the authors of SPSS Statistical Algorithms when they published that book, making that package more useful.

(b) Statistical computing.

Thisted's work in statistical computing supported by this contract has resulted in several reports. With Thomas Sager [4], he has developed theoretical results for maximum likelihood estimation in isotonic modal regression problems, together with fast computer algorithms for computing the estimators. This work is applicable in examining populations whose modes are known to satisfy certain order restrictions. In addition, a fast memory-efficient algorithm has been constructed for computing the distribution for a family of rank tests. The Kolmogorov-Smirnov test is one member of this family. This work was reported in an invited talk at a meeting of the Institute of Mathematical Statistics [5].

In conjunction with his work on Monte Carlo methods and random number generation, Thisted has published a detailed review of Statistical Methods for Digital Computers [6], and is continuing work on a survey of random number generation, including methods designed specifically for small computers and programmable calculators.

A FORTRAN program has been written which does the computations associated with the diagnostic checks and data assessment procedures developed for ridge regression estimators discussed in the next section.

(c) Regression methods.

In this area, Thisted has principally studied ridge regression and related procedures for improving on least-squares estimation in the presence of ill-conditioned data (multicollinearity). Proponents

of ridge estimators have claimed, on the basis of simulation studies, that ridge estimators have mean squared-error (MSE) everywhere less than that of least squares estimators. Work in [7] shows that certain generalized ridge estimators suggested in the literature cannot be better than least squares everywhere. Moreover, these procedures have larger MSE than least squares under circumstances likely to be encountered in practice. Consequently, such procedures are not recommended for routine use.

A second report [8] deals with "ordinary" ridge regression estimators, which are used most frequently in practice. The paper contains a thorough investigation of multicollinearity and the manner in which biased estimators in general, and ridge estimators in particular, address the problems induced by multicollinearity. The bias introduced by these estimators can be thought of as additional information supplementing the data. From this viewpoint one can point out conditions under which this additional information is correct in some sense, so that ridge estimators are likely to improve on least squares.

In an invited talk [9] at the Special Topics Meeting on Regression of the Institute of Mathematical Statistics, Thisted reported on recent advances in ridge regression theory and the degree to which ridge estimators can be inferior to least squares. When multicollinearity is severe, ridge estimators can have substantially larger MSE.

Other work on ridge regression, although not supported by this contract, has stemmed directly from or has profited by the work outlined above. Thisted and Carl Morris have developed theoretical results which can be applied to most adaptive ordinary ridge estimators. They present necessary and sufficient conditions for such estimators to be everywhere better than least squares. Thisted has also critiqued the ridge trace and other methods for doing data analysis arising from ridge estimators.

References

- [1] Thisted, R. (1979). User documentation and control language I: Evaluation and comparison of statistical computer packages. Computers & Education 3 pp. 135-141.
- [2] Thisted, R. (1979). Teaching statistical computing using computer packages (with Discussion). American Statistician 33 pp. 27-35.
- [3] Thisted, R. (1979). Comments on SPSS Statistical Algorithms. Proceedings of the 3d Issue conference.
- [4] Thisted, R. and Sager, T. (1979). Maximum likelihood estimation of isotonic modal regression. Technical Report #81, Department of Statistics, University of Chicago.
- [5] Thisted, R. (1978). Statistical algorithms and dynamic programming. Invited presentation given at the Eastern Regional Meeting of the Institute of Mathematical Statistics, New Brunswick, New Jersey.
- [6] Thisted, R. (1979). Review of Statistical Methods for Digital Computers. Computing Reviews 20 pp. 309-312,328,348,358.
- [7] Thisted, R. (1978). On generalized ridge regression. Technical Report #57, Department of Statistics, University of Chicago.
- [8] Thisted, R. (1978). Multicollinearity, Information, and Ridge Regression. Technical Report #66, Department of Statistics, University of Chicago.
- [9] Thisted, R. (1979). Perspectives on ridge regression. Invited presentation given at the IMS Special Topics Meeting on Regression, Boulder, Colorado.

VIII Comparison of several populations.

Investigator: James A. Koziol.

(a) Multiple comparisons of K populations differing by scale parameters (with N. Reid).

A typical problem arising in life testing situations is that of determining whether K populations differ from one another by a shift in scale: for example, upon observing independent samples of exponentially distributed random variates from K populations, one might wish to detect multiplicative treatment effects. Given samples from two populations with underlying exponential distributions, the statistic introduced by Savage (1956) yields the locally most powerful rank test of the null hypothesis that the two populations are identically distributed, against the Lehmann alternative that they differ by a scale parameter. This two sample test may be extended to a test against differences in scale when $K > 2$ samples are given, using a general form of K sample rank tests described by Hajek and Sidak (1967, Section III.4.1).

When it is of less importance to test whether the K populations are identically distributed than to determine which populations differ from one another on an individual basis, a multiple comparisons procedure is preferable to a single global test. Koziol and Reid have developed two nonparametric methods of comparing K populations. Under the first procedure, all the observations in the combined K samples are jointly ranked, and any two populations are compared with average rank scores as determined from the Hajek-Sidak K sample generalization of the Savage test. Under the second procedure, any two populations are compared with rank scores from the pairwise ranking of their respective samples, this pairwise ranking excluding observations from the other $K-2$ samples. It is shown that the two methods of comparing all pairs of populations have equivalent asymptotic properties, including the same asymptotic relative Pitman efficiency. This work appears in [1] and [2].

(b) Multiple comparisons among K samples subject to unequal patterns of censorship.

A common problem is the analysis of incomplete survival data, where complete survival data on various individuals or components is lacking, due to their disappearance (censoring) from the study at some time. Among the statistical procedures for testing equality of the underlying survival distributions of two populations when observations are arbitrarily censored is Gehan's modifications of the Wilcoxon two sample rank test. Breslow (1970) has proposed a generalization of Gehan's procedure for testing equality of K continuous survival distributions when observations in each of the $K > 1$ samples may be subject to arbitrary censorship. However a single global test of whether the K populations are equivalent is somewhat uninformative; it would be of greater interest to determine which, if any, of various treatments enhance survival. Koziol has addressed this problem, and derived a multiple comparisons procedure for arbitrarily censored populations. This work appears in [3].

- (c) A test for bivariate symmetry based on the empirical distribution function.

Consider the hypothesis of bivariate symmetry

$$H: F(x,y) = F(y,x) \text{ for all } x,y \quad (1)$$

concerning an arbitrary absolutely continuous distribution F . Hollander (1971) proposed a procedure for testing (1) based on the bivariate empirical distribution function of the observations. Hollander's test statistic is not invariant under transformations of the coordinates by the same monotone function, hence is not distribution free. In view of this Hollander determined the permutation distribution of his statistic conditional upon the observed sample, and thereupon proposed a conditional distribution-free test of bivariate symmetry.

Suppose now that the null hypothesis (1) is restricted to

$$H: F(x,y) = G(x)G(y) \text{ for all } x,y \quad (2)$$

where G is an arbitrary absolutely continuous distribution function. (2) being a subhypothesis of (1), it follows from Hollander's work that his statistic is consistent for (2). Moreover the asymptotic distribution of his statistic is distribution-free under (2), and so is of interest. Koziol has investigated this asymptotic distribution, and shown it to be that of a infinite weighted sum H of chi-square variates each with one degree of freedom. Critical values of H are obtained by numerical inversion of its characteristic function. A general formulation for the asymptotic power in testing (2) against a sequence of local alternatives is given. A particular case, a version of the sphericity hypothesis for bivariate normal observations, is examined in greater detail, with a numerical power comparison of Hollander's statistic with the best parametric test being made. This work appears in [4].

- (d) Sequential testing of the equality of two survival distributions. (with J. Petkau).

The problem of testing the equality of two survival distributions is confounded when the subjects, be they mechanical components or human beings, are lost from the study for one reason or another before its completion. Koziol and his coauthor Petkau have investigated a sequential version of the modified Savage statistic used in testing equality of two survival distributions when the observations are singly censored. The asymptotic distribution of the statistic is found, and other asymptotic properties are studied. In particular, when the underlying distributions are exponential, it is shown that the asymptotic power of this sequential procedure differs little from that of the optimal procedure which utilizes all the available information, but that the sequential procedure affords a substantial time savings in arriving at a decision. Comparison is made between the modified Savage statistic and the modified Mantel-Haenszel statistic recently proposed by Muenz, Green, and Byar in *Biometrika* (1977) for the same hypothesis testing problem. The MGB statistic seems less

useful, in that its critical values are available only when the sample sizes are equal and small; since the MGB method to determine critical values for larger sample sizes is in question, the behavior of the MGB statistic when samples sizes are large or unequal is unknown. This work appears in [5].

References

- [1] Koziol, J. and Reid, N. (1978). Multiple comparisons of K populations differing by scale parameters. Mathematische Operationsforschung und Statistik, Ser. Statist 9 179-184.
- [2] Koziol, J. and Reid, N. (1977). On the asymptotic equivalence of two ranking methods for k-sample linear rank statistics. Annals of Statistics 5 pp. 1099-1106
- [3] Koziol, J. and Reid, N. (1977). Multiple comparisons among K samples subject to unequal patterns of censorship. Communications in Statistics A6 (12) pp. 1149-1164
- [4] Koziol, J. (1979). A test for bivariate symmetry based on the empirical distribution function. Communications in Statistics. A8 (3) pp. 207-221.
- [5] Koziol, J. and Petkau, J. (1978) Sequential testing of the equality of two survival distributions using the modified Savage statistic. Biometrika. 65 pp. 615-623

XI. Unbiasedness in Multivariate Statistics

Investigator: Michael D. Perlman

Perlman has carried out several investigations into the unbiasedness of multivariate test statistics. In [1], a question first addressed by Pitman in 1939 is resolved, namely the unbiasedness of Bartlett's modification of the likelihood ratio test (LRT) for testing equality of the covariance matrices of three or more multivariate normal populations. Similar techniques led to the surprising result that for the related problem of testing equality of both mean vectors and covariance matrices simultaneously, it is the unmodified LRT, not the Bartlett-like modification often proposed, which is unbiased. The same methods were also applied in [4] to study the power functions of the LRT's for testing problems involving the complex multivariate normal distribution. On the other hand, quite different methods (essentially involving a multivariate generalization of the notion of monotone likelihood ratio) were developed in [2] to establish the unbiasedness of the class of all admissible invariant tests for the multivariate analysis of variance (MANOVA), for testing independence of two sets of variates, and other problems. The methods in [2] were also applied in [4]. Finally, a summary of recent progress in these areas is given in [3].

References

- [1] Perlman, M.D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. Annals of Statistics 8 pp. 247-263.
- [2] Perlman, M.D. and Olkin, I. (1978). Unbiasedness of invariant tests for MANOVA and other multivariate problems. Technical Report #70, Department of Statistics, University of Chicago. To appear in Annals of Statistics.
- [3] Perlman, M.D. (1980). Unbiasedness of multivariate tests: recent results. Proceedings of the Fifth International Symposium on Multivariate Analysis. North-Holland. pp 413-432.
- [4] Perlman, M.D. and Anderson, S.A. (1979). Two testing problems relating the real and complex multivariate normal distributions. Technical Report #95, Department of Statistics, University of Chicago. Submitted to Journal of Multivariate Analysis.

C. CONSULTATION AND COOPERATIVE RESEARCH WITH OTHER SCIENTISTS AT
THE UNIVERSITY OF CHICAGO AND ARGONNE NATIONAL LABORATORIES

Through its consultation program the Department of Statistics participates in other research projects at the University of Chicago and Argonne National Laboratories of potential interest to DOE. Coordinators of the consultation program were: for 1975-76, Roderick Little; for 1976-77, Alan Izenman; for 1977-78, Paul Meier, Kent Bailey, Michael Friedman; for 1978-79, Peter McCullagh. All members of the Department of Statistics, as well as its graduate students, participate in the consultation to varying degrees. Some activities of the program for 1975-79 are briefly reviewed.

1. Statistical Analysis of Moon Rocks. [Dr. John Morgan, Department of Chemistry]. The proportion of various metals in rock samples collected by Apollo missions should provide information about meteoritic activity on the surface of the moon. On the basis of 74 samples, Dr. Morgan and colleagues thought that the rocks could be grouped into 6 groups or clusters, and they wanted more objective evidence that the groupings existed. Factor analysis had been tried with little success. Little rejected factor analysis and cluster analysis as unsuitable for the small sample size. With the help of graduate student Stella Machado, a principal component analysis was carried out on a selection of 38 substantial rock samples, and the rock samples plotted out in the coordinates of the first two principal components, which captured most of the variability in the measurements. The results provided a useful picture of the data, but provided very little evidence of grouping.

2. Statistical Estimation of Particle Counts from an Orbiting Spacecraft. Mr. John Jump (Laboratory for Astrophysics and Space Research) consulted on an indirect method for counting atomic particles arriving at an orbiting spacecraft, as a substitute for a malfunctioning direct recording device. An estimated rate for the particle in question was obtained by assuming a Poisson distribution for arrivals, and Mr. Jump had been given a formula for the standard error of the estimate which he mistrusted. His doubts proved well founded, for Little's analysis, using propagation of error and the theory of conditional probability, indicated a missing term in the formula. Mr. Jump was also advised about how to analyze the data for fluctuations in the particle rates over time.

3. Visit of the Department of Statistics to Argonne: 13 January 1976. Six faculty members and seven graduate students from the Department of Statistics visited Argonne on 13 January 1976 at the invitation of Dr. James Pool, Associate Director of the Applied Mathematics Division of Argonne National Laboratories. Statistical problems arising in research projects at Argonne were presented in talks by Dr. Tom Baldwin (Energy and Environmental Systems Division), Dr. George Sacher (Biological and Medical Research Division), Dr. Stephen Spigarelli (Radiological and Environmental Research Division). These discussions were general in nature, and emphasized the need for statistical expertise in the analysis of a wide range of data. Dr. David Raske (Material Sciences Division) presented a particular problem in some detail, and received

statistical advice.

Dr. Raske described data collected from fatigue analysis, where a metal is subjected to repeated flexing under various physical conditions, and the number of cycles to breakage (N) recorded. The aim is to find a prediction equation for N as a function of four independent variables: T , the temperature of metal, S , the log strain range, R , the log strain rate, and H , the hold time. From a sample of 430 observations, Dr. Raske had fitted linear regressions of various functions of N on polynomial functions of T , S , R , and H , and had arrived at a prediction equation, expressing

$\frac{1}{\log_{10} N}$ as a linear function of 18 polynomials in T , S , R , and H . The equation explained 96% of the variance of the dependent variable, and produced residuals which appeared to be normally distributed with constant variance. Discussion by Professors Wallace, Shelby Haberman and Izenman centered on the choice of dependent variables, and on the need to constrain the stepwise regression procedures to retain to sensible and interpretable groups of terms usually by hierarchical procedures. In 1977 contact relating to this problem was re-established and is discussed in a later item.

4. Visit of Drs. James Pool, James Lyness, and party from Argonne Applied Math Division to the Department of Statistics, University of Chicago: 18 February 1976. Dr. James Lyness gave a consulting seminar entitled "Evaluating Quadrature Routines." The performance profile of various automatic quadrature routines were compared over a family $f(\lambda)$ of integrands. For a specified nominal level of accuracy ϵ_{req} set as a parameter of the routine, the attained accuracy $\epsilon_{\text{act}}(\lambda)$ and the number of function evaluations $v(\lambda)$ were measured over the family of integrands to provide two performance profiles. Since for automatic routines this often has a profile "like the Chicago skyline", comparisons of routines are highly dependent on λ . Large numbers of λ in the given range are generated by a random number generator, and the random variables $\epsilon_{\text{act}}(u)$ and $v(u)$ studied, where u is a uniform variable over the range of λ .

Dr. Lyness had found that 5-10% of the time an automatic routine can fail, in the sense that ϵ_{act} was much larger than the required accuracy ϵ_{req} . He presented ways of calibrating ϵ_{req} to be 95% sure that ϵ_{act} is less than a given accuracy, ϵ_0 . He also estimated the mean number of function evaluations, $\bar{v}_{.90}(\epsilon_0)$, required at the calibrated value of ϵ_{req} . This was used to compare the performance of different quadrature routines on the given problem.

Dr. Lyness' talk stimulated a lively discussion over a period of two hours. Active participation for the Department of Statistics included Professors Wallace, Little, Haberman, Izenman, Wichura, Zabell, and Perlman. Among the points discussed were the following: (a) An economical presentation of the data, in view of the apparently Gaussian nature of the distribution of $\epsilon_{\text{act}}(u)$; (b) further study of ϵ_{act} and v as joint random variables, for example, by plots of ϵ_{act} against v to explore dependence, possibly leading to other measures such as ϵ_{act}/v , the attained accuracy per function evaluation. (c) The use of the statistical theory of estimating a distribution function to suggest the

number of λ -values needed to estimate the distribution of ϵ_{act} to a given accuracy.

5. Statistical Analysis of Nuclear Radiation Detector Counts. (Mr. Michael Neumann, a physicist, Franklin McLean Cancer Research Institute and Mr. John Shepherd, the Institute for Computer Research) were interested in testing the sensitivity of a new type of gamma-ray detector; such detectors are commonly used in medical diagnostic studies of the metabolic and physiologic functions of various organs, including the brain. Mr. Neumann had designed and constructed a detector which, on the basis of some preliminary studies, appeared to have extremely favorable performance characteristics when compared to similar detectors currently in use.

Several experimental runs were carried out in Mr. Neumann's laboratory using a single point source of radiation (germanium 68, typically preferred for brain imaging) and a detector of the kind constructed in that laboratory. These runs were of various durations, with the majority 100 secs. Interest centered first on showing that for an experimental run of given duration, the intervals between successively recorded events followed an exponential distribution. A frequency distribution (or, histogram) of interval lengths for each run was computed by Mr. Shepherd. However, a limitation of the counting mechanism meant that no interval of length smaller than 20 μ secs could be recorded; thus, even though events less than 20 μ secs apart had been observed by the detector, the counter was not able to record such intervals, and so they did not appear in any of the histograms. Furthermore, the histograms that were produced did not use equal 'bins' or class-intervals. Instead, to bring in a very long tail, 20 μ sec-length bins were judiciously grouped into 'slices' which were then used as new class-intervals for the resulting histogram; these 'slices' were of differing lengths, depending on how many bins were combined to make the 'slice'. The frequency corresponding to a particular 'slice' was computed as the average of the frequencies of the bins comprising that 'slice'. These histograms, however, did not appear to be of exponential shape. Subsequent investigations by Izenman showed that incorrect readings were being taken due to a faulty setup of the equipment used in the experimental runs.

Some suggestions were then made by Izenman on how to analyze the mass of data that was potentially available. One was to design an experiment that would take into account the following possible sources of variation that could occur in a given experimental run: (1) total duration of the run (e.g. 50 secs, 100 secs, 200 secs, ..., 1000 secs); (2) voltage used (e.g. 0-3000 volts, with special emphasis on the range 1500-2500 volts); (3) position of detector relative to the source. A further suggestion was that a printout of all individual interval-lengths for each run in the order in which they occurred should be made and would be very useful as a diagnostic tool. If such details were readily available then modelling of the stream of events could be undertaken. This stream, in theory, should be a stationary Poisson process, and if details on the occurrence times of events for each run were available, then a spectrum analysis of the point process generated could be used to check the sensitivity of the detector.

6. Assessing Low Level Radiation. Dr. George Sacher of Argonne sent to the Department of Statistics a proposed project entitled "A Statistical Evaluation of Alternatives in Initiating a Long-Term Experiment Using Dogs," written by S. A. Tyler, G. A. Sacher, and W. P. Norris. He asked that it be reviewed from the point of view of statistical design, etc. The underlying problem was that of assessing the long-range effects of low-level radiation in lesser animals. The proposal contained two key elements, namely (a) the identification of breakpoints in the regression of log death-rate on log dose, and (b) the effective allocation of animals in the assessment of the regression slope (with a highly restricted choice sets for 3 doses, i.e., 80, 40, and 40; or 40, 80, and 40; or 40, 40, and 80). The reviewers, Professors Paul Meier, Shelby Haberman, and Joan Chmiel questioned whether the two-slope model of point (a) above was really satisfactory, and whether some continuous-slope model might be comparably informative and more tractable. They also suggested an alternative approach via the Cox regression model [D. R. Cox, J. Royal Statistical Soc. Series B, (1972), p. 187] in which an unknown underlying life-distribution, or equivalently, death-rate function, is assumed, and the effect of radiation is modelled as a multiplier of that function. They were also concerned about how the two segments of the two-slope model (each fitted by least squares) are to be separated and suggested a literature search of papers on regression involving breakpoints. Concerning point (b), Meier thought that a reasonable thing to do would be to seek a performance measure which reflected the desired discriminatory power, and thereby determine its value for each of the three alternatives. Sacher replied that the paper had been revised on the basis of these comments.

7. Plant Resistance to Insect Predation. An ecologist interested in the defense plants use to resist insect predators postulated that the defense may depend on the "life strategy", or ecological niche of the plants, specifically, whether the plant is a permanent or transient inhabitant. Permanent plant species might be more apt to be toxic, while those with shorter life cycles would depend upon lack of information on the insects' part. A sample of species were categorized in this way as permanent or transient, and a sample of trees of each species were studied by marking a sample of leaves, and recording the percent of the leaf area eaten in a 100-day period. The chief statistical issue is the one of arriving at a representative measure of plant success in view of the possible confounding of different defense mechanisms: if a leaf is untouched, the success could be due to either toxicity or lack of discovery, while if partially eaten, success must be due to toxicity. Several measures and approaches are being considered, including dealing with the "zeroes" in a separate analysis. The random effects model rears its head here and the methodological question arises whether to transform at the lowest level (leaf) or to aggregate first.

8. Relationship of DNA Content of Neuroblastoma to Tumor Malignancy. This study aims to find out whether benign and malignant tumors of the nerves of small children can be distinguished on the basis of a normal or abnormal cell DNA content. Some 35-40 tumor tissue slides were prepared--one per case--and read under an optical densitometer. The integrated optical density in a cell is presumed to be proportional to DNA content. One statistical concern was the proper allowance to be made for different machine calibrations in different

sessions. Standard reference cells were always measured along with the 30 or so study cells per case. Graphical comparison of independent readings of the same cells on different calibrations showed the machine effect to be a change of scale, not a translation. There is some interest in the sources of measurement error (broadly defined): reader error, cell selection, and machine drifts. However the main issue is what composite measure of the distribution of DNA content is to be used diagnostically, if indeed there is any effect on survival. The last two questions are still being studied.

9. Changes in Behavior of Animals. First-order Markov chains are being used to model animal behavior at successive points in time according to a given set of categories. Of interest is the development of statistical methods to determine at what points, if any, in the sequence of behaviors, there occurs a change in the hypothetical transition matrix generating the Markov chain. A promising procedure has been devised which incorporates chi-square testing of empirical transition matrices constructed from non-overlapping segments of the chain. This procedure has been tried on simulated data, and found to perform well. One very interesting problem is the distribution of the combined chi-square value under the null hypothesis of no changes in the transition matrix.

10. Fatigue Testing of Stainless Steel. Interaction in 1977 with Argonne National Laboratory centered in a sequence of consultations in summer and fall between David Wallace, and Dwight Diercks and David Raske of the Materials Science Division relating to the reanalysis of their extensive data on elevated temperature, strain-controlled fatigue data on stainless steel discussed in item 3 above. Graduate student Daniel Keenan spent part of the summer at Argonne working on the reanalysis. Major issues centered on allowing for systematic effects of the testing laboratory and of the heat treatment of the steels and of choosing reexpressions of the main design variables that improved additivity of stepwise regression packages on polynomial terms, to a careful buildup of models with rationally added terms led to a fitted surface that was interpretable and yet gave reasonable fits. The data remain subject to some erratic behavior probably due to the lack of control in the original collection of data, but much improved analyses resulted.

11. Effect of Electric Shock Stimulus on Trans-membrane Voltage. Marc Walton, a graduate student interested in the response rate of heart tissue to electric shock stimulus, has designed an experiment where the changing voltage across a cell membrane is measured and traced on an oscilloscope. The voltage trace, however, is of relatively less importance than the rate of change of voltage and the second derivative of the voltage trace. Electronic differentiators introduce so much noise that the resulting traces are virtually worthless. McCullagh, and graduate student Mark Jaeger are trying to model the shape of the voltage trace by a smooth function thereby eliminating high frequency noise from the system. The fitted function can then be differentiated analytically eliminating the need for electronic differentiation. The main statistical problems are (i) that the voltage trace has an asymmetric ogive shape (ii) the oscilloscope trace must first be digitized manually and (iii) not all points have similar variability.

D. PUBLICATIONS RECEIVING FULL OR PARTIAL DOE/ERDA SUPPORT:

15 JUNE 1975 -- 14 JUNE 1979

FACULTY

R. R. Bahadur

Bahadur, R.R. and Zabell, S. (1979). Large deviations of the sample mean in general vector spaces. Annals of Probability. 7 pp. 587-621.

Bahadur, R.R. (1980). On combining independent tests of significance. In preparation.

Bahadur, R.R., Koziol, J.A., and Wallace, D.L. (1980). Comparing test efficiencies -- leading asymptotic terms can be misleading. In preparation.

Michael D. Perlman

Perlman, M.D. and Eaton, M.L. (1976). Generating $O(n)$ with reflections. Technical Report #29, Department of Statistics, University of Chicago. (Sent 9/76)

Perlman, M.D. (1977). A note on the matrix-variate F distribution. Indian Journal of Statistics 39 Series A . pp. 290-298.

Perlman, M.D. and Eaton, M.L. (1977). Reflection groups, generalized Schur functions, and the geometry of majorization. Annals of Probability 5 pp. 829-860.

Perlman, M.D. and Koziol, J. (1978). Combining independent non-central chi-square tests. Journal of the American Statistical Association 73 pp. 753-763.

Perlman, M.D. and Olkin, I. (1978). Unbiasedness of invariant tests for MANOVA and other multivariate problems. Technical Report #70, Department of Statistics, University of Chicago. To appear in Annals of Statistics.

✓ Perlman, M.D. and Anderson, S.A. (1979). Two testing problems relating the real and complex multivariate normal distributions. Technical Report #95, Department of Statistics. University of Chicago. Submitted to Journal of Multivariate Analysis.

✓ Perlman, M.D. Koziol, J., and Rasmussen (1980). Combining independent non-central F-tests. In preparation.

Perlman, M.D. and Marden, J. (1980). Combining independent chi-square and F tests: admissibility and complete class theorems. Technical Report, Department of Mathematics, Urbana.

✓ Perlman, M.D. and Larntz, K. (1980). A simple procedure for testing equality of correlation matrices. In preparation.

✓ Perlman, M.D. and Marden J. (1980). Invariant tests for means with covariates. Annals of Statistics 8 pp. 25-63.

Perlman, M.D. and Lin, S. P. (1980). A Monte Carlo comparison of four methods for estimating a covariance matrix. Technical Report, Department of Mathematical Sciences, Memphis State University. Submitted to Journal of the American Statistical Association.

Perlman, M.D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. Annals of Statistics 8 pp. 247-263.

✓ Perlman, M.D. (1980). Unbiasedness of multivariate tests: recent results. Proceedings of the Fifth International Symposium on Multivariate Analysis. North-Holland. pp 413-432.

✓ Ronald Thisted

Thisted, R. (1978). On generalized ridge regression. Technical Report #57, Department of Statistics, University of Chicago. (Sent 9/78)

✓ Thisted, R. (1978). Multicollinearity, Information, and Ridge Regression. Technical Report #66, Department of Statistics, University of Chicago. (Sent 9/78)

Thisted, R. (1979). User documentation and control language I: Evaluation and comparison of statistical computer packages. Computers & Education 3 pp. 135-141.

Thisted, R. (1979). Teaching statistical computing using computer packages (with Discussion). American Statistician 33 pp. 27-35.

✓ Thisted, R. (1979). Comments on SPSS Statistical Algorithms. Proceedings of the 3d Issue conference.

Thisted, R. and Sager, T. (1979). Maximum likelihood estimation of isotonic modal regression. Technical Report #81, Department of Statistics, University of Chicago.

Thisted, R. (1979). Review of Statistical Methods for Digital Computers. Computing Reviews 20 pp. 309-312, 328, 348, 358.

David L. Wallace

Wallace, D.L. (1977). Pearson curves as asymptotic approximations. Technical Report #42, Department of Statistics, University of Chicago. (Sent 9/77)

Wallace, D.L. (1979). SNAP User's Manual, October 1979, 116 pages, University of Chicago Computation Center On-line Documentation.

Wallace, D.L. (1979). Plotting on the TEKTRONIX in DEC-SNAP, SNAP Memorandum Number 10, July 79 (Revised).

Wallace, D.L., and Wichura, M.J. (1979). PSAVE and PEND: Routines to save data in SNAP's SAVE format, SNAP memorandum Number 11, September 1979.

Wallace, D.L. (1980). Interpolating FORTRAN computations within SNAP, SNAP memorandum Number 12, July 1980

Wallace, D.L. (1980). The Behrens-Fisher and Fieller Creasy problems. In Fienberg, S.E. and D.V. Hinkley (Editors), R. A. Fisher: An Appreciation, Lecture Notes in Statistics, 1, pp. 119-147. New York: Springer-Verlag.

Wallace, D.L. (1980). Analytic approximations to distributions of largest eigenvalues. In preparation.

Wallace, D.L. (1980). Gamma probability plotting can be simple. In preparation.

Wallace, D.L. (1980). Sampling cumulants of multivariate generalized k-statistics for inhomogeneous sampling. In preparation.

Bahadur, R.R., Koziol, J.A., and Wallace, D.L. (1980). Comparing test efficiencies -- leading asymptotic terms can be misleading. In preparation.

Michael J. Wichura

Wichura, M.J. (1979). On the probability of a symmetric stable process crossing a bottom outer boundary. Z. Wahrscheinlichkeits-theorie und verwandte Gebiete. 47 pp. 195-203.

- Wallace, D.L., and Wichura, M.J. (1979). PSAVE and PEND: Routines to save data in SNAP's SAVE format, SNAP memorandum Number 11, September 1979.
- Wichura, M.J. (1980). A computer program for the numerical evaluation of first passage distributions. In preparation.
- Wichura, M.J. (1980). Boundary crossing probabilities for the Wiener process. In preparation.
- Wichura, M.J. (1980). Boundary crossing probabilities for extremal processes. In preparation.

✓Sandy L. Zabell

- Zabell, S.L. and Izenman, A. (1976). Babies and the Blackout: A time series analysis of New York city birth data. Technical Report #38, Department of Statistics, University of Chicago. (Sent 10/76)
- Zabell, S.L. (1976). Arbuthnot, Heberden, and the Bills of Mortality. Technical Report #40, Department of Statistics, University of Chicago. (Sent 5/77)
- Zabell, S.L. (1979). Continuous versions of regular conditional distributions. Annals of Probability 7 pp. 159-165

DOCTORAL DISSERTATIONS AND DERIVED PUBLICATIONS

H. Brown (Research Advisor: David L. Wallace)

- Brown, H. (1980). Missing values in factor analysis. In preparation.

S. P. Lin (Research Advisor: Michael D. Perlman)

- Lin, S. P. (1978). An improved procedure for the estimation of a correlation matrix. Technical Report #69, Department of Statistics, University of Chicago. (Sent 9/78)
- Lin, S.P. and Perlman, M.D. (1980). A Monte Carlo comparison of four methods for estimating a covariance matrix. Technical Report, Department of Mathematical Sciences, Memphis State University. Submitted to Journal of the American Statistical Association.

J. Marden (Research Advisor: Michael D. Perlman)

- Marden, J. (1978). Complete class theorems for invariant tests in multivariate analysis. Technical report #52, Department of Statistics, University of Chicago. (Sent 9/78)
- Perlman, M.D. and Marden J. (1980). Invariant tests for means with covariates. Annals of Statistics 8 pp. 25-63.

✓ Perlman, M.D. and Marden, J. (1980). Combining independent chi-square and F tests: admissibility and complete class theorems. Technical Report, Department of Mathematics, University of Illinois at Urbana.

Marden, J. (1980). Invariant tests for means with covariates. Technical Report, Department of Mathematics, University of Illinois at Urbana. Submitted to Annals of Statistics.

S. G. Machado (Research Advisor: David L. Wallace)

Machado, S.G. (1976). On transformations of multivariate data. Technical Report #23, Department of Statistics, University of Chicago. (Sent 9/76)

A. G. de Wet (Research Advisor: David L. Wallace)

de Wet, A.G. (1977). Linear functional and structural relationships. Technical Report #54, Department of Statistics, University of Chicago. (Sent 9/78)

VISITING FACULTY

James Alexander Koziol

Koziol, J. and Reid, N. (1977). On the asymptotic equivalence of two ranking methods for k-sample linear rank statistics. Annals of Statistics 5 pp. 1099-1106

✓ Koziol, J. and Reid, N. (1977). Multiple comparisons among K samples subject to unequal patterns of censorship. Communications in Statistics A6 (12) pp. 1149-1164

Koziol, J. and Reid, N. (1978). Multiple comparisons of K populations differing by scale parameters. Mathematische Operationsforschung und Statistik, Ser. Statistik 9 179-184. (TR sent 9/76)

✓ Koziol, J. and Petkau, J. (1978). Sequential testing of the equality of two survival distributions using the modified Savage statistic. Biometrika. 65 pp. 615-623

✓ Koziol, J. (1979). A test for bivariate symmetry based on the empirical distribution function. Communications in Statistics. A8 (3) pp. 207-221.

✓ Bahadur, R.R., Koziol, J.A., and Wallace, D.L. (1980). Comparing test efficiencies -- leading asymptotic terms can be misleading. In preparation.

RECEIVED BY TIC

SEP 9 1980