(2)

ARO Report 90-2

PROCEEDINGS OF THE THIRTY-FIFTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Approved for public release; distribution unlimited.
The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

U.S. ARMY RESEARCH OFFICE

Report No. 90-2

August 1990

PROCEEDINGS OF THE THIRTY-FIFTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

TRADOC Test and Experimentation Command
Experimentation Center (TEC)
Fort Ord, California

HELD AT

Monterey Beach Hotel
Monterey, California
13-20 October 1989

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina

FOREWORD

The Thirty-Fifth Conference on the Design of Experiments in Army Research, Development and Testing had as its host the TRADOC Test and Experimentation Command, Experimentation Center (TEC), Fort Ord, California. This conference was planned for 18-20 October 1989, and was held in the Monterey Beach Hotel, Monterey, CA. The earthquake on 17 October prevented several of the speakers from attending this meeting; and while the power was off, problems arose for many of the speakers. Dr. Marion Bryson, Director of TEC, served as local host and conference coordinator. He and members of his staff are to be commended for supplying innovative and immediate solutions to many problems associated with the quake. Without their support the conference would never have succeeded.

The Army Mathematics Steering Committee (AMSC) is the sponsor of the Conference on the Design of Experiments. Members of this committee would like to thank D. Hue McCoy, TRADOC Analysis Command, for organizing the Special Session on "Statistical Issues Related to Combat Modeling." The speakers were Hue McCoy, Bill Baker (BRL), and Eugene Dutoit (Infantry School). This session achieved its purpose of stimulating a dialogue between combat modelers and the statistical community. The AMSC members feel that the addresses by the principal speakers, as well as the contributed papers by Army and academic personnel, also stimulated the interchange of ideas among the scientists attending this meeting. Noted below is the list of invited speakers selected by the Program Committee:

Speaker and Affiliation	Title of Address
Professor Robert Bechhofer Cornell University	An Appraisal of Several Multistage Selection Procedures
Professor William J. Conover Texas Tech University	Latin Hypercube Sampling, a Way of Saving Computer Runs
Professor Gary Koch University of North Carolina at Chapel Hill	An Overview of Statistical Methods for Categorical Data
Professor David W. Scott Rice University	Statistical Data Analysis

Another event associated with each of these conferences is a two-day tutorial. This year, Ronald Hocking of Texas A&M University presented a tutorial entitled "Analysis of Linear Models with Unbalanced Data." It was held two days before the start of the conference and was conducted in the TEC Protocol Building at Fort Ord.

As the master of ceremonies at the banquet and the recipient of the Wilks Award last year, Dr. Marion Bryson had the honor of announcing the winner of the ninth U.S. Army Wilks Award, Professor Boyd Harshbarger. He was selected because of his research endeavors, his promotional activities for Army applications, his unending supply of speakers for these conferences, and his help in numerous ways to carry the Army forward in many important statistical areas. Because of ill health, Professor Harshbarger was unable to attend the conference. Dr. Douglas Tang, representing the Army statistical community, accepted the award on his behalf.

Members of the Army Mathematics Steering Committee would like to thank the members of the Program Committee for guiding this scientific conference, and to also thank the Mathematical Sciences Division of the Army Research Office for preparing the proceedings of these meetings.

PROGRAM COMMITTEE

Carl Bates
Eugene Dutoit
Douglas Tang
Robert Burge
Hue McCoy
Malcolm Taylor
Henry Tingey
Francis Dressel
Carl Russell
Jerry Thomas

TABLE OF CONTENTS*

*This Table of Contents contains only the papers that are published in this technical manual.

TABLE OF CONTENTS (continued)

AGENDA

THE THIRTY-FIFTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT, AND TESTING

18-20 October 1989

Host: TRADOC Test and Experimentation Command
Experimentation Center (TEC)
Fort Ord, California 93941-7000
Marion R. Bryson, Director

Location: Monterey Beach Hotel
2600 Sand Dunes Drive
Monterey, California 93940

Wednesday, 18 October 1989

0730 - 0900 **REGISTRATION**

0915 - 0930 **CALLING THE CONFERENCE TO ORDER:**

Marion R. Bryson, Director
TRADOC Test and Experimentation Command
Experimentation Center (TEC)

WELCOMING REMARKS

0930 - 1200 **GENERAL SESSION I**

Chairperson: Marion R. Bryson, TRADOC Test and Experimentation Command, Experimentation Center

0930 - 1030 **KEYNOTE ADDRESS:**

AN APPRAISAL OF SEVERAL MULTISTAGE SELECTION PROCEDURES
Robert Bechhofer, Cornell University

1030 - 1100 **BREAK**

1100 - 1200 STATISTICAL DATA ANALYSIS
David W. Scott, Rice University

1200 - 1330 **LUNCH**

1330 - 1500 **CLINICAL SESSION A**

Chairperson: Barry Bodt, U.S. Army Ballistic Research Laboratory

Panelists: William J. Conover, Texas Tech University
Jayaram Sethuraman, Florida State University
Nozer Singpurwalla, George Washington University

HAS VARIABILITY BEEN REDUCED?
Gary Aasheim, U.S. Army Armament, Munitions and Chemical Command

WHICH DISTRIBUTION APPLIES?
Gary Aasheim, U.S. Army Armament, Munitions and Chemical Command

******

1330 - 1500 **TECHNICAL SESSION 1**

Chairperson: Francis Dressel, U.S. Army Research Office

MODELING DEPENDENCE INDUCED BY COMMON ENVIRONMENTS
Mark A. Youngren, U.S. Army Concepts Analysis Agency

EVALUATION OF DESERT CAMOUFLAGE UNIFORMS BY GROUND OBSERVERS
George Anitole, Ronald L. Johnson, U.S. Army Belvoir Research, Development and Engineering Center, and Christopher Neubert, U.S. Army Materiel Command

ELIMINATING CALCULUS DEPENDENCY IN THE DERIVATION OF DODGE'S u
Richard M. Brugger, U.S. Army Armament, Munitions and Chemical Command

HOW SHOULD ERROR ESTIMATES OF FIXED CAMERA CALIBRATION CONSTANTS BE COMPUTED?
William S. Agee and Andrew C. Ellingson, U.S. Army White Sands Missile Range

1500 - 1530 **BREAK**

Wednesday (Continued)

1530 - 1710 TECHNICAL SESSION 2

Chairperson: Malcolm Taylor, U.S. Army Ballistic Research Laboratory

PROMOTING STATISTICAL LITERACY AND INTERACTION OF RESEARCHERS AND STATISTICIANS
Emanuel Parzen, Texas A&M University

BAYESIAN INFERENCE FOR NONHOMOGENEOUS POISSON POINT PROCESSES USING EXPERT OPINION AND DATA
Nozer D. Singpurwalla, George Washington University

RANDOM MAPPINGS
Bernard Harris, University of Wisconsin-Madison

Thursday, 19 October 1989

0815 - 0945 APPLICATION SESSION

Chairperson: Carl Bates, U.S. Army Concepts Analysis Agency

HANDLING UNCERTAINTY IN EXPECTED VALUE MODELS
Mark A. Youngren, U.S. Army Concepts Analysis Agency

APPLICATION AND CALIBRATION OF A STOCHASTIC C^3 COMBAT MODEL FOR OUTER-AIR AND INNER-AIR BATTLES
Izhak Rubin, University of California at Los Angeles and Israel Mayk, U.S. Army Communications and Electronics Command

LOADING AND MATERIAL PROPERTY UNCERTAINTIES IN FINITE ELEMENT ANALYSES FOR ORTHOPAEDICS
Shirish Chinchalkar and D. L. Taylor, Cornell University

******

0815 - 0945 TECHNICAL SESSION 3

Chairperson: Jock O. Grynovicki, U.S. Army Human Engineering Laboratory

NONNEGATIVE ESTIMATION OF VARIANCE COMPONENTS IN MIXED LINEAR MODELS WITH TWO VARIANCE COMPONENTS I
Thomas Mathew, University of Maryland

NONNEGATIVE ESTIMATION OF VARIANCE COMPONENTS IN MIXED LINEAR MODELS WITH TWO VARIANCE COMPONENTS II
Bimal Kumar Sinha, University of Maryland

NONPARAMETRIC INFERENCE FOR IMPERFECT REPAIR MODELS
Jayaram Sethuraman, Myles Hollander, and Brett Presnell, Florida State University

0945 - 1015 BREAK

1015 - 1200 CLINICAL SESSION B

Chairperson: Carl Russell, U.S. Army Operational Test and Evaluation Agency

Panelists: Robert Bechhofer, Cornell University
Bernard Harris, University of Wisconsin
Emanuel Parzen, Texas A&M University

APPLICATION OF A COMPOSITE DESIGN TO TEST A COMBAT SIMULATION MODEL
Carl B. Bates, U.S. Army Concepts Analysis Agency

APPLICATION OF RESPONSE SURFACE METHOD TO RANDOM VIBRATION
Mircea Grigoriu, Cornell University

******

1015 - 1200 TECHNICAL SESSION 4

Chairperson: John Robert Burge, Walter Reed Army Institute of Research

DISTRIBUTION THEORY FOR VARIANCE COMPONENT ESTIMATION DIAGNOSTICS
Jock O. Grynovicki, U.S. Army Human Engineering Laboratory and John W. Green, University of Delaware

TECHNICAL SESSION 4 (Continued)

NUMERICAL ESTIMATION OF THE PARAMETERS OF THE SOURCE DENSITY FUNCTION
Charles E. Hall, Jr., U.S. Army Missile Command

THE HUNTER PROBLEM IN A RANDOM FIELD OF OBSCURING ELEMENTS
Shelemyahn Zacks and M. Yadin, State University of New York at Binghamton

1200 - 1330 LUNCH

1330 - 1530 SPECIAL SESSION

Chairperson: D. Hue McCoy, U.S. Army TRADOC Analysis Command

STATISTICAL ISSUES RELATED TO COMBAT MODELING
D. Hue McCoy, U.S. Army TRADOC Analysis Command

A NONPARAMETRIC APPROACH TO THE VALIDATION OF STOCHASTIC SIMULATION MODELS
William E. Baker and Malcolm S. Taylor, U.S. Army Ballistic Research Laboratory

SMALL SAMPLE TESTS OF SIGNIFICANCE IN SUPPORT OF COMBAT MODELING
Eugene Dutoit, U.S. Army Infantry School

1530 - 1600 BREAK

1600 - 1700 GENERAL SESSION II

Chairperson: Gerald R. Andersen, U.S. Army Research Office

LATIN HYPERCUBE SAMPLING, A WAY OF SAVING COMPUTER RUNS
William J. Conover, Texas Tech University

1830 - 1930 CASH BAR

1930 - 2130 BANQUET AND PRESENTATION OF WILKS AWARD

Friday, 21 October 1989

0815 - 0945 **TECHNICAL SESSION 5**

Chairperson: William S. Agee, White Sands Missile Range

THE VARIANCE OF THE INTEGRATED PROCUREMENT PROBLEM VARIABLE - A FRESH APPROACH
Barnard H. Bissinger, Pennsylvania State University

GRAPHICAL TOOLS FOR EXPERIMENT DESIGN
Russell R. Barton, Cornell University

MONTE CARLO SURFACE APPROXIMATION USING ORTHOGONAL FUNCTIONS
Peter W. Glynn and Donald L. Iglehart, Stanford University

0945 - 1015 **BREAK**

1015 - 1145 **GENERAL SESSION III**

Chairperson: Douglas B. Tang, Walter Reed Army Institute of Research; Chairman of the AMSC Subcommittee on Probability and Statistics

OPEN MEETING OF THE STATISTICS AND PROBABILITY SUBCOMMITTEE OF THE ARMY MATHEMATICS STEERING COMMITTEE

AN OVERVIEW OF STATISTICAL METHODS FOR CATEGORICAL DATA ANALYSIS
Gary Koch, University of North Carolina at Chapel Hill

ADJOURN

PROGRAM COMMITTEE

Carl Bates
Eugene Dutoit
Douglas Tang
Robert Burge
Hue McCoy
Malcolm Taylor
Henry Tingey
Francis Dressel
Carl Russell
Jerry Thomas

STATISTICAL DATA ANALYSIS: HOW FAR WILL COMPUTER GRAPHICS TAKE US?

David W. Scott
Department of Statistics
Rice University
P.O. Box 1892
Houston, Texas 77251-1892

ABSTRACT. In this paper we survey the directions researchers are following in statistical graphics. Hardware support for animation and of color is expanding rapidly while price is at least decreasing. While a fairly optimistic scenario can be drawn, the most correct statement we can make about the future of graphics and statistical computing is that the uncertainity has never been greater. Potential obstacles towards effective use of computer graphics are discussed, particularly in the academic setting. Strategies to break these bottlenecks will be suggested. Otherwise excess CPU cycles may remain so.

1. INTRODUCTION. Each year at the annual meeting of the National Computer Graphics Association, a gala dinner is held at which the winners of various computer graphics contests are presented. As the winning computer-generated images and videos are presented, with bumble bees darting among flowers and pool balls reflecting the images of a futuristic shiny room, one is overwhelmed by the shear raw power and impact of the presentation. There is not (yet) a category for statistical presentation, but one senses this is not out of the question.

The impact of modern computer graphics on statistical education and practice has not yet been great. Eddy et al. in a recent article in *Statistical Sciences* have attempted to describe future computing needs and trends, and graphics is an important part of the overall picture. The average statistician retains a small collection of typical images that are recycled over and over: scatter diagrams including residual plots, frequency curves such as histograms, curve fits such as regression lines, elliptical contours of normal densities including principal components; the list is surprisingly small. Far more emphasis is given to tables: summary statistics tables, chi-squared tables, analysis of variance tables, tables of percentiles, and spreadsheets. This follows the natural inclination of statisticians to present a parsimonious summary of an incidence of data analysis: choose a powerful model well-studied in the literature, estimate parameters and determine significance, and present results summarizing the model in tabular and sometimes graphical forms. Image processing, animation, rotation are all very unparsimonious statistical tools.

Historically, technology has affected the relative importance of these forms. Early data analysts such as John Graunt and William Petty favored tabular presentation, after all, paper was a dear commodity. William Playfair showed the array of graphical presentation of business data was worth the paper. Computation was expensive, and the human effort required for creating effective graphs was relatively cost-effective. Karl Pearson began the trend towards testing and tabular presentation, but devoted much energy to graphs in the form of frequency curves. Fisher and others accelerated the tabular form with analysis of variance and maximum likelihood, which emphasizes parametric analysis over the more graphical nonparametric analysis. The emphasis was on mathematical statistics. The rapid increase in number crunching ability spawned the creation of statistical packages, with largely numerical output. Graphics was not ignored in such packages (certainly not in the past few years), but the quality was relatively low and options limited. Quality graphics output is still much more expensive than computing, but the absolute price of both has decreased so dramatically that we are seeing an explosion of interest in graphical statistics. Truly impressive packages for personal computers are available and SAS and SPSS have provided similar capabilities for mainframes. Separately, many non-statistical companies provide software for presentational graphics, aimed at business markets. ISCOL is one example, but such quality products cost even academic workers many thousands of dollars.

2. CURRENT IMPACT OF COMPUTER GRAPHICS. How strong has the impact of computer graphics been on the statistical community? To look at many journals and statistical textbooks, you would be hard pressed to detect any revolution. In its fourth edition, Hogg and Craig's classical textbook on mathematical statistics contains only five figures! The Journal of the American Statistical Association is showing the change, but in unexpected ways. Roughly half of the papers contain only tables. Those with figures contain more figures than papers ten years ago, but ironically the quality is poorer. Ten years ago artwork was professionally drawn (if only approximating truth). Many figures today are drawn by PC's, which are acceptable but clearly inferior in presentation quality and impact of their professional cousins. But the cost is so much less that we accept substandard quality. The very recent increase in laser graphical output partially justifies the premature switch to PC graphics.

The long and short of it is that we are within five years of everyone having the ability to produce very high quality two-dimensional graphics virtually without cost. In other words, we have succeeded in automating the kinds of graphs William Playfair drew 200 years ago.

3. NEW DIRECTIONS IN COMPUTER GRAPHICS. The emphasis of this paper is on how much farther will computer graphics take statistics? Why is there a trend towards newer graphical presentations? Graphics is at odds with classical statistics because graphics is nonparsimonious. A graph cannot be neatly summarized or reduced to a few key coefficients and p-values. Graphs demand close scrutiny and invite speculation and interpretation, something hardly ever seen in parametric analyses. But the fundamental distinguishing feature is that graphs are subjective, imprecise, manipulative, yet powerful. One novel multivariate graph is the Chernoff face. An entire conference in 1978 was devoted to evaluating the subjective aspects of this technique, in particular, coping with the almost infinite possible alternative constructions for *individual* datasets. There is no consensus whether it is a serious statistical tool. The discipline of statistics attempts to be very precise about its imprecision, and many statisticians do not find graphs precise enough to serve as the analysis, preferring tables and statistics.

Yet the whole new technology of computer graphics and enhanced graphics chips has opened up the possibility of a new generation of presentation graphics. More statisticians are focusing their research effort in this area, and are represented by the new ASA section called statistical graphics. The concerns about limitations of the old style graphics are even more critical in the new style of graphics. The key additional features are color, solids rendering, translucency, and animation; the Pixar machine is the state-of-the-art for all of these features. If we consider the exploratory graphical tools for high dimensional data, we see that an important part of data analysis is luck. For the higher the dimension, the smaller the fraction of data that can be "explored" in a given amount of time. Thus different workers examining the same multivariate data will probably see *disjoint* parts of it - quite in contrast to a parametric world using principal components. Even the order in which the data are examined can be a factor, given the inevitable fatigue. Some research is already under way to help automate the searching process (reminds me of the computer science project to automate the game Rogue, called rogomatic). But real objections have been made about this imprecise form of data analysis. The use of color excludes those who are color blind. The use of stereo viewing techniques is maddeningly unsuccessful for a large percentage of professionals. Each new subjective element increases the power of the data analysis but decreases the reliability and widespread usefulness of these techniques. Publishing is virtually impossible, until CD-ROM publishing is available. A nonexhaustive list of projects includes: projection pursuit (Tukey, Friedman, Stuetzle); animated scatter plots (Tukey, Huber, Donoho); exploratory methods (Tukey and Tukey); density estimation (Scott, Thompson, Tarter); glyphs and stereo (Carr and Nicholson); grand tours (Buja and Asimov); programming languages (Becker, Chambers, Donoho, Huber); programming environments (McDonald).

4. MANAGING THE FUTURE. But enough about how hard it all will be and how unappreciated it all may be. Are we going to be able to sustain research in novel statistical graphics? As an engineering undergraduate in 1968, I used to wait in line to use a Wang time-sharing calculator terminal (it actually could do the transcendental functions to twelve significant digits!). Once

we began doing our number crunching through programming languages, we could accept and track the new computing resources with almost no overhead. So in the past fifteen years, I have written Fortran (and PL/I) programs on as many types of hardware. The only overhead was learning a new editor, a few system commands, and the faster and bigger machine was immediately increasing productivity and opening new horizons. There is still a bit more of that to be had. With the workstations now available, we have finally obtained the luxury of wasting a huge fraction of CPU cycles. This is of course a correct state of affairs given the relative cost of faculty time. Idle CPU seconds are costly only in terms of maintenance; idle graphics workstations cannot yet be justified as maintenance costs are very high.

But we must face two developments. The first is parallel computing. The second is graphics. Statisticians can probably make the most effective use of parallel computers than any single group of researchers, because much of our computing involves very loosely coupled computation such as Monte Carlo simulation. Numerical analysts, on the other hand, face tightly coupled computation which provides real gains only in rather specific situations. Theoretical limits exist to performance in tightly coupled systems, no matter how many parallel processors are available. But all that aside, to effectively use hypercube or other parallel architectures is not a straightforward exercise. It is even worse than having to give up your favorite programming language and return to assembler. Serious allocation of time and other supporting resources must be made at this time. One reaction is that it is not worth the effort and just to wait until some computer scientist writes an incredible parallel compiler that takes non-parallel code and optimizes into parallel environments. (Not too likely in my opinion. Gene Golub at Stanford in a comment after a lecture by John Rice lamented that there weren't enough numerical analysts to go around to try and make parallel algorithms for each differential equation and hardware configuration.)

Graphics presents the same challenge. With more modest effort, one can produce useful pictures on a PC or graphics terminal of the William Playfair variety. Playing with the color tables can be fun. Choosing the specific 256 colors from the 16,777,216 choices can be a bit frustrating. Graphics chips have helped enormously, putting frequently used graphical transformations into hardware and supporting animation. The interface with these chips is at about the same level as other graphics commands, almost at the assembler level, pixel by pixel. Some systems are available at the command level to avoid this, but the convenience eventually becomes the limitation, both in functionality and performance. At a somewhat lower level, graphics standards have appeared, such as CORE and GKS. But any commercial outfit will admit that the advantages of portability are outweighed by the benefits of performance allowed by assembler programming. But most academics are satisfied by "prototype" systems rather than commercial performance.

My observation is that with graphics systems it is very difficult to build upon previous work. Each new generation of hardware demands a complete new attack. As the graduate students who did the previous system disappear, the next generation of students have a more difficult task getting up to speed. For the better hardware often has many more capabilities, so reproducing the previous system often much harder. Therefore, less time is available for extending the previous system and actually less research gets done. This is a bit overdrawn, but accurately reflects what has happened over the past fifteen years. At Berkeley, a biostatistical researcher developed a analysis and graphical system on some IBM hardware that he nursed for eight years beyond its supported lifetime, before finally biting the bullet and updating hardware. At Rice and Stanford and other places, graduate students who worked on very specialized hardware and produced very useful systems, graduated and went away. What was left was a collection of faculty who had directed the research but who did not have the time to actually program the system, maintain it, or even fully understand it. Thus the next generation of graduate student basically found it impossible to effectively use the machines. Maintenance costs and down-time were significant as the expensive hardware aged, and using the previous student's system frustrating (and not research). The apparent time to start new and create a wholly new system was determined too risky, since rumors that the machine might be sold (since no one was using it) began to circulate. The traditionally successful faculty/graduate student relationship was found wanting. The need for continuity implied the need for a new type of person in the picture (nontraditional), the staff support group. These persons can usually be recruited from recent graduates by offering post-docs,

research positions, and other positions not commonly found in statistics groups. Thus there is a need to restructure research personnel to continue this work. The systems are too complex for individual faculty to manage (much less to retrain unproductive faculty). Fewer and fewer graduate students are able to master the complexities of these systems in the few years available and make real contributions. Those who can leave quickly, leaving behind a serious void in continuity, rendering expensive equipment unusable almost overnight. These statistics and computer science wizards are not well-recognized as doing valid statistical research worthy of tenure track (as opposed to statistical computing). The result is inability to do the desired research, which necessarily includes extensive systems development. We seem to be moving towards the system used by sciences, many post-docs per faculty member as well as support staff to provide full-time research effort and continuity of systems expertise and support, something that cannot be even partially satisfied by faculty and students alone. Unfortunately, the job market is so strong in statistics as opposed to these other areas that it will be very difficult to build up new centers and move towards the big research lab model.

This will be a rather traumatic trend. It is well-known that using programmers greatly reduces output (due to decreased reliability of code and less intimate knowledge of the problem) and decreases hands-on experimentation that leads to new developments, but senior faculty time can not usually be allocated significantly for this purpose. Debugging purely graphical systems is extraordinarily difficult. Dr. Banchoff at Brown University reports that Roger Penrose found a bug in a four-dimensional hidden-line removal algorithm by simply watching it perform. Testing will be an enormous headache and problem. Everything looks so pretty when the output is graphics. Difficult to be critical. We have watched computer science departments try and manage very large development projects. Statistical researchers will have to pay attention to how these efforts have been organized and managed. Statisticians seem to be a bit impatient and more satisfied with prototypes of systems than is healthy for the profession.

Another approach has been to move to novel computing environments that hold the promise of improved user productivity and portability. The LISP machines fall into this category.

At Battelle Labs in Richland, Washington, Wes Nicholson and Dan Carr have pioneered research into the use of glyphs and stereo viewing for data analysis. In 1983 they invited a distinguished panel of statisticians and computer scientists to review and criticize their progress. It is clear from the reprinted papers and discussion that the visitors could not decide what was "fundamental research" and what was merely "systems development." This lack of a clear understanding of the joint roles of these activities has hindered the professional development of many young computer-bound statisticians.

5. CONCLUSIONS. We asked the question of how far will computer graphics take us? The answer is a long way, but not with the current research structure. Graphics requires as much support as supercomputing or parallel architectures, but may not get it directly. Many of the sciences and engineering departments have received adequate laboratory resources and statistics must be added to the list. The need for and trend towards graphics can not be altered, but we can work on improving presentation quality and effectiveness, such as Bill Cleveland (1985) and others have been attempting to evaluate. Statisticians have contributed much to the burgeoning field of "scientific visualization," but it is computer scientists who have dominated the funding in the field. A closer working relationship to the fields of application is already occurring but more should be expected. Finally, examples of figures shown in the original talk may be found in the references below.

6. ACKNOWLEDGMENTS. This work was supported in part by grants from the ONR and the ARO, N00014-90-J-1176 and DAAL-03-88-K-0131, respectively. The original talk was greatly varied to cope with the unusual circumstances surrounding the great California earthquake the preceding day. I wish to thank the organizers for their professionalism under such stress.

7. REFERENCES.

Asimov, D. (1985), "The Grand Tour," *SIAM J. Sci. Statist. Comput.*, 6:128-143.

Barnett, Vic (1981), *Interpreting Multivariate Data*, Wiley, New York.

Becker, R.A. and Chambers, J.M. (1984), *S: An Interactive Environment for Data Analysis and Graphics*, Wadsworth, Belmont, CA.

Becker, R.A., Chambers, J.M., and Wilks, A.R. (1988), *The New S Language*, Wadsworth Brooks/Cole, Pacific Grove, CA.

Carr, D.B. and Nicholson, W.L. (1984), "Graphical Interactive Tools for Multiple 2- and 3-Dimensional Scatterplots," *Proceedings of NCGA*, Anaheim, CA.

Cleveland, W.S. (1985), *The Elements of Graphing Data*, Wadsworth, Monterey, CA.

Eddy, W. (1986), "Computers in Statistical Research," Workshop on the Use of Computers in Statistical Research, *Statistical Sciences* 1:419-453.

Foley, James D., and Van Dam, Andries (1982), *Fundamentals of Interactive Computer Graphics*, Addison-Wesley, Reading, MA.

Friedman, J.H., McDonald, J.A., and Stuetzle, W. (1982), "An Introduction to Real-Time Graphical Techniques for Analyzing Multivariate Data," *Proceedings of NCGA*.

Hall, D.L. (1983), "ALDS 1982 Panel Review," Pacific Northwest Laboratory Report PNL-SA-11178, Richland, WA.

Haerdle, W. and Scott, D.W. (1989), "Smoothing in Low and High Dimensions by Weighted Averaging Using Rounded Points," Rice Technical Report 88-16.

Jee, R. (1985), "A Study of Projection Pursuit Methods," Unpublished Dissertation, Rice University.

Lorensen, W.E. and Cline, H.E. (1987), "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," *Computer Graphics*, 21:163-169.

Scott, D.W. (1983), "Nonparametric Probability Density Estimation for Data Analysis in Several Dimensions," Proceedings of the Twenty-Eighth Conference on the Design of Experiments in Army Research Development and Testing, pp. 387-397.

Scott, D.W. (1984), "Multivariate Density Function Representation," Proceedings of the Fifth Annual National Computer Graphics Association Conference, Volume II, pp. 794-801.

Scott, D.W. (1986), "Data Analysis in 3 and 4 Dimensions With Nonparametric Density Estimation," in *Statistical Image Processing and Graphics*, E.J. Wegman and D. DePriest, Eds., Marcel Dekker, New York, pp. 291-305.

Scott, D.W. (1987), "Software for Univariate and Bivariate Averaged Shifted Histograms," Rice Technical Report 311-87-1.

Scott, D.W. (1990), "Statistics in Motion: Where Is It Going?" Proceedings of Statistical Graphics Section, to appear.

Scott, D.W. and Hall, M.R. (1989), "Interactive Multivariate Density Estimation in the S Language," Proceedings of the 20th Interface of Computer Science and Statistics, American Statistical Association, Alexandria, Virginia

Scott, D.W. and Thompson, J.R. (1983), "Probability Density Estimation in Higher Dimensions," *Computer Science and Statistics: Proceedings of the 15th Symposium on the Interface*, pp. 173-197, North Holland, New York.

Terrell, G.R. and Scott, D.W. (1990), "On Adaptive Estimation of Multivariate Densities," Working Paper.

Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.

Tukey, John W. (1977), *Exploratory Data Analysis*, Reading, MA.

Wegman, E. (1990), "Hyperdimensional Data Analysis using Parallel Coordinates," *J. American Statistical Assoc.*, to appear.

Wegman, E.J. and DePriest, D.J. (1986), *Statistical Image Processing and Graphics*, Marcel Dekker, New York.

EVALUATION OF DESERT CAMOUFLAGE UNIFORMS BY GROUND OBSERVERS

GEORGE ANITOLE AND RONALD L. JOHNSON
U.S. ARMY BELVOIR RESEARCH, DEVELOPMENT AND ENGINEERING CENTER
FORT BELVOIR, VIRGINIA 22060-5606

CHRISTOPHER J. NEUBERT
U.S. ARMY MATERIEL COMMAND
ALEXANDRIA, VIRGINIA 22333-0001

ABSTRACT

The standard U.S. Army desert camouflage uniform appears dark against U.S. and Saudi Arabian desert backgrounds. Prototype uniforms were developed and evaluated in the desert Southwest in 1986. Test results led to further evaluation, in 1987, of seven new uniforms, plus the standard uniform. Uniforms were shown in all possible pairs, at ten sites, to U.S. Marine Corps and Fort Belvoir personnel, who served as ground observers. The uniforms were judged on their ability to blend with the background. The best of each pair was independently selected. An analysis of variance and Duncan's Multiple-Range Test statistics were performed. It was determined for most sites, and across all sites, that three new uniforms were significantly ($\alpha \leq 0.05$) best in blending with the background.

1.0 SECTION 1 - INTRODUCTION

The standard U.S. Army desert camouflage uniform is made in a pattern consisting of six colors. The predominant color areas are tan, khaki, light brown, and dark brown. Small light-brown areas outlined in black are scattered throughout the other color areas. This uniform was taken to Saudi Arabia in 1980, and viewed against multiple desert backgrounds. In all cases the uniform appeared dark and did not blend well with any of the observed desert backgrounds. This information was given to counter-surveillance personnel at Natick RD&E Center, MA. A series of seven prototype desert uniforms was then made and given to Fort Belvoir for a desert evaluation in 1986. Analysis of this data[1/] identified uniforms 4, 5, and 6 as being the most effective in terms of blending with the U.S. desert test sites investigated.

Using the additional test information collected by Belvoir as a basis, Natick then developed uniforms 8, 9, 10, and 11 for further evaluation. These uniforms, along with uniforms 4, 5, and 6 and the standard U.S. Army uniform, identified as uniform 1, were evaluated in the U.S. desert Southwest in 1987. The quantitative analysis of their ability to blend with various Southwest desert backgrounds is the subject of this report.

2.0 SECTION 2 - PROCEDURE

2.1 Test Uniforms

A total of eight camouflage uniforms were evaluated. The following is a description of each uniform:

- Uniform #1--Standard U. S. Army Desert Day Camouflage Pattern
  A six-color pattern now in use by the U.S. military consisting of the colors Light Tan 379*, Tan 380*, Light Brown 381*, Dark Brown 382*, Black 383*, and Khaki 384*.

- Uniform #4
  A three-color pattern of Light Tan 379*, Khaki 384*, and Light Brown 381*.

♦ Uniform #5
A three-color pattern of Light Tan 379*, Tan 380*, and Khaki 384*.

♦ Uniform #6
A three-color pattern of Desert Tan 459*, Khaki 384*, and Light Brown 381*.

♦ Uniform #8
A solid-color uniform of Tan 380*.

♦ Uniform #9
A solid-color uniform of Khaki 384*.

♦ Uniform #10
A three-color pattern of Khaki 384*, brown** and sand**.

♦ Uniform #11
A two-color pattern of clay** and Khaki 384*.

*Natick numerical color designations
**No numbers assigned

2.2 Test Sites

A total of ten sites were selected for the study. All the desert sites contained sparse vegetation similar to that found in areas of interest in the Middle East. The soil ranged in color from a light buff/tan to gray and dark brown, and represented a good cross-sectional spectrum of different-colored desert backgrounds. The order of the ten sites as they will appear throughout this study is seen in Table 1.

Table 1
Site Order Identification

Site #	Color	Location
1	Buff	Yuma Sand Dunes, AZ
2	Light Gray	Ogilby Road, Tumco, CA
3	Very Light Tan	Yuma Proving Grounds, AZ
4	Dark Beige Tan	Anza Borrego State Park, CA
5	Light Tan	Tank Trail, 29 Palms, CA
6	Dark Tan	Salton Sea, CA
7	Beige Tan	Anza Borrego State Park, CA
8	Light Beige Tan	Anza Borrego State Park, CA
9	Tan	Jean Dry Lake Bed, NV
10	Gray Tan	Rt. 15, Baker, CA

2.3 Test Subjects

The test subjects consisted of U.S. Marine Corps enlisted men from Camp Pendleton, CA, and civilians from the U.S. Army Natick Research, Development, and Engineering Center, Natick, MA, and the U.S. Army Belvoir Research, Development, and Engineering Center, Fort Belvoir, VA. A maximum of 15 observers to a minimum of 10 observers were used at each test site. All subjects had at least a corrected visual acuity of 20/30 and normal color vision.

2.4 Data Generation

The eight uniforms were viewed, individually, in all possible pairs (28). The viewing distance from the subject to each pair of uniforms was about 25 meters. The observers were

told to select the one uniform from each pair that best matched or blended with the surrounding background in terms of color. The observers were instructed to discount shrubbery if present. This instruction was necessary, because of the very sparse shrubbery in the deserts of the Middle East when compared with the U.S. desert Southwest. The mean preference with associated standard error, 95% confidence intervals, analysis of variance, and Duncan's Multiple-Range[2/] were calculated for all sites, and averaged across all ten sites. The higher the mean preference, the more preferred the colors were rated by the ground observers as blending with the desert background.

3.0 SECTION 3 - RESULTS

The camouflage uniforms were evaluated at each of the ten sites to determine which colors best blended with the desert environment. Section 2.4 describes how the data was generated for all sites, and when averaged across all sites. Table 2 shows the uniforms that best blended with each site and when averaged across all sites.

Table 2
Summary of the Best Desert Uniforms for Each Site In Ability to Blend with the Background

	Uniforms							
	1	4	5	6	8	9	10	11
Site 1			X		X			
Site 2		X	X		X			X
Site 3		X	X		X		X	X
Site 4		X	X					X
Site 5				X			X	
Site 6		X	X		X		X	
Site 7		X	X		X			
Site 8			X		X			
Site 9		X	X		X			
Site 10			X				X	X
Across All Sites		X	X		X			

The statistical results of each site for the above best camouflage uniforms will not be included, because they would be too voluminous to present in these proceedings. This data is available upon request from the U.S. Army Belvoir Research, Development and Engineering Center, ATTN: STRBE-JDA, Fort Belvoir, VA 22060. Table 3 contains the mean preference with associated standard error and 95% confidence interval for the ability of the desert uniforms to blend with the background, when averaged across all sites. Figure 1 is the graphic display of Table 3. Table 4 is the analysis of variance performed to determine if there are significant differences between the various camouflage uniforms in their ability to blend with the desert backgrounds. Table 5 identifies which uniforms differ from each other through the Duncan's Multiple-Range Test.

Table 3
Mean Preference Rating for Desert Background Blend and 95-Percent Confidence Intervals (Across All Sites)

Uniform	N	Mean	Standard Error	95% Confidence Interval Lower Limit		Upper Limit
1	116	0.8190	0.0761	0.6683	to	0.9696
4	116	4.3966	0.1266	4.1458	to	4.6473
5	116	4.7845	0.1340	4.5190	to	5.0500
6	116	2.5345	0.1725	2.1928	to	2.8761
8	116	4.5000	0.1197	4.2630	to	4.7370
9	116	0.9397	0.0902	0.7610	to	1.1184
10	116	3.9655	0.1278	3.7124	to	4.2187
11	116	3.6466	0.1878	3.2745	to	4.0186

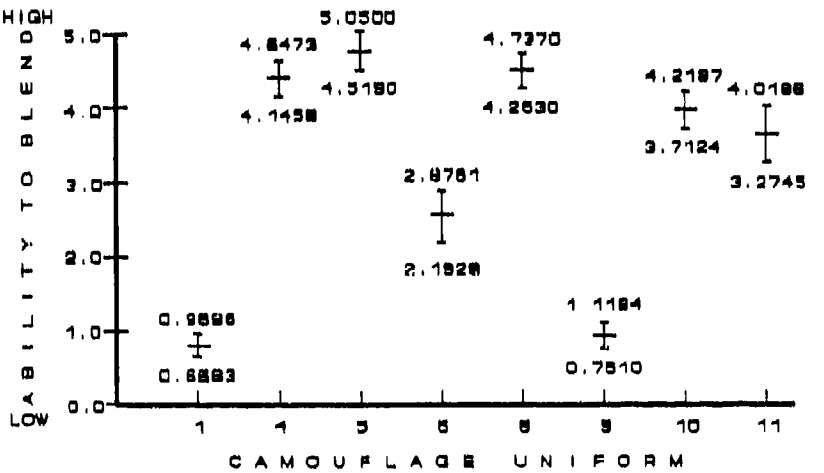


Figure 1
Desert Camouflage Uniform Ability to Blend with the Desert Background, Means, and 95-Percent Confidence Intervals (Across All Sites)

Table 4
Analysis of Variance for the Ability of the Camouflage Uniforms to Blend with the Desert Background (Across All Sites)

Source	Degrees of Freedom	Sum of Squares	Mean Square	F-Test	Level
Uniforms	7	2046.1379	292.3054	140.4009	0.0000*
Error	920	1915.3793	2.0819		
Total	927	3961.5172			

Bartlett's Test for Homogeneous Variance
Number Degrees of Freedom = 7
F = 19.23 Significance Level = 0.000**

*Significant at α less than 0.001 level

Table 4 indicates that there are significant differences in the ability of the camouflage uniforms to blend with the desert background. The Bartlett's Test indicates that the variance for each uniform is not homogeneous, i.e., significantly different, so they are not necessarily from the same population.

Table 5
Duncan's Multiple-Range Test
for All Sites Combined, Daylight

BEST	1	UNIFORM 4 4.3966	UNIFORM 8 4.5000	UNIFORM 5 4.7845
	2	UNIFORM 11 3.6466	UNIFORM 10 3.9655	
	3	UNIFORM 6 2.5345		
WORST	4	UNIFORM 1 0.8190	UNIFORM 9 0.9397	

4.0 SECTION 4 - DISCUSSION

A review of the data for sites 1-10, and for all sites combined, shows that camouflage uniforms 4, 5, and 8 were the most effective in blending with the desert terrain. These uniforms had mean blending values of 4.3966, 4.7845, and 4.5000 respectively (Tables 3 and 5). With the exception of site 5 (Table 2), where camouflage uniforms 6 and 10 were judged as best blending with the desert background, uniforms 4, 5, and 8 had at least one member among those that blended best with the desert background. The overall mean-blending values for the uniforms do not differ significantly from each other (Table 5 and Figure 1). Additional review of the data indicates that the standard camouflage uniform (#1) and uniform 9 had the worst blend with the desert background, when averaged across all sites.

The data for this study appears fairly clean; however, one large and pressing caveat must be taken into consideration, before any final decision on desert uniforms is made. The uniform tests conducted so far have been in the U.S. desert Southwest. Any future conflicts in which a desert camouflage uniform will be used by U.S. forces will, in all probability, be in the Middle East. These deserts tend to be lighter and more tan than the grayer desert of the United States. They also have much less vegetation. The best camouflage uniforms from this study should be evaluated in the areas of interest in the Middle East for final determination as to color blend with the background. The resulting data may necessitate color modifications of the uniforms to ensure that the best possible blend with the deserts of interest is achieved.

5.0 SECTION 5 - SUMMARY AND CONCLUSIONS

A total of eight camouflage uniforms were evaluated as to their ability to blend with desert backgrounds in the U.S. desert Southwest. Ten sites were used. The uniforms were viewed in all possible pairs (28), and with the one selected from each pair that blended best with the background. The results of this evaluation produced the following conclusions:

a. Camouflage uniforms 4, 5, and 8 blended best with the U.S. desert backgrounds.

b. Standard camouflage uniform 1 and prototype uniform 9 were the least effective in blending with the U.S. desert backgrounds.

c. An additional desert camouflage evaluation should be conducted in the Middle East, to ensure that the best uniform is selected for the U.S. military.

REFERENCES

1. Letter Report, STRBE-JDS to U.S. Army Natick RD&E Center, "Development of an Effective Desert Camouflage Pattern/Color for Uniforms," 11 July 86.

2. Natrella, Mary G., Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.

HAS VARIABILITY BEEN REDUCED?

Gary Aasheim
U.S. Army Armament, Munitions and Chemical Command
Product Assurance and Test Directorate
Tool and Equipment/Aircraft Armament Branch
Rock Island, Illinois 61299-6000

Often changes are made in measuring methods and in production methods with at best, only checks to determine whether or not the changes affected variability. After a change is made, a natural question is - Did the change affect measurement precision or product uniformity?

I am not aware of an established method for analyzing before and after sample results to answer that question for all situations. Of course, if the before and after change samples are from the same population, the standard F-test can be used.

But sometimes the before-change samples are from one set of populations and the after-change samples are from a different set of populations.

One method for dealing with this situation is to compare the pooled before change variance with the pooled after change variance using an F-test. However, if one or both sets of populations are heteroscedastic, this method seems to be of marginal soundness. What are some possible approaches for dealing with this latter situation?

WHICH DISTRIBUTION APPLIES?

Gary Aasheim
U.S. Army Armament, Munitions and Chemical Command
Product Assurance and Test Directorate
Tool and Equipment/Aircraft Armament Branch
Rock Island, Illinois 61299-6000

1. Faced with the questions - do the sample measurements support the customer's belief that a given dimensional requirement was not met to the degree required by the contract, and, if not, what dimensional requirements could be met to the required degree? - a co-worker of mine took the 60 sets of 20 readings (see below) and checked for normality by:

a. transforming the readings in each set by dividing each difference, reading minus set sample average, by the set sample standard deviation.

b. treating the 1200 transformed readings as a single sample of 1200.

c. finding the average, standard deviation, skewness and kurtosis of the transformed readings, plus the standard deviations of the latter two statistics based upon the assumption that the 1200 readings were from a normally distributed population.

d. breaking the transformed readings by size into 26 groups and running a chi-square goodness-of-fit test where the expected values were based upon the normal distribution.

2. Two considerations drove the transforming and pooling efforts above. First, running 60 tests for normality would have taken more time and work than the approach taken. Second, when my co-worker gained an initial acquaintance with the data by computing sample averages and standard deviations and by counting readings outside the dimensional requirements, he did not spot any obviously atypical readings and, so, felt that an assumption of a single underlying statistical distribution with different parameters for different populations was reasonable.

3. Is there a better approach than that used by my co-worker?

LOT DATA TO ACCOMPANY 'WHICH DISTRIBUTION APPLIES?' PAPER FOR 35TH DESIGN OF EXPERIMENTS CONFERENCE

L O T N U M B E R

67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
62	64	77	63	58	63	82	61	81	62	90	68	72	65	62	65	64	63
64	63	88	64	53	73	88	60	85	61	84	64	66	64	54	55	60	69
61	64	84	63	60	63	94	71	87	60	85	62	68	61	62	54	73	69
64	62	78	64	68	65	101	76	83	55	99	65	58	63	64	55	65	69
69	61	80	61	63	63	107	61	82	60	82	65	62	63	62	59	65	70
67	63	88	65	63	65	86	60	88	62	80	70	69	61	65	61	69	74
74	64	86	62	68	62	92	66	71	59	88	57	67	64	62	59	66	70
72	64	74	59	68	69	89	66	85	60	77	60	74	65	59	63	63	70
74	61	75	63	62	62	89	58	84	59	83	60	65	62	61	63	59	70
72	62	78	66	51	63	70	58	83	62	83	59	62	60	60	63	66	73
64	68	85	65	74	66	94	63	73	65	74	59	68	63	66	62	68	70
64	64	82	63	62	66	82	63	89	58	77	60	64	65	60	65	64	63
66	61	87	66	58	62	87	65	86	62	83	61	65	65	57	73	61	71
59	61	81	65	64	62	96	69	81	62	78	60	66	60	58	58	65	69
67	68	81	63	54	67	95	62	100	58	80	62	61	66	69	65	71	66
67	66	87	66	68	65	95	68	78	58	81	58	71	62	60	74	77	69
58	65	85	62	64	68	87	65	91	62	85	58	66	65	62	69	62	70
60	62	83	68	56	67	90	64	85	60	81	68	67	59	62	64	52	66
59	63	87	65	66	62	90	63	86	64	72	61	62	59	63	60	61	76
62	64	75	59	59	63	86	61	81	62	75	64	58	67	57	62	63	72

L O T N U M B E R

85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102
63	71	65	63	61	70	60	64	58	66	63	64	61	67	73	65	56	66
61	70	64	71	68	58	57	66	57	67	54	63	62	67	63	64	62	68
56	66	58	68	60	67	64	68	59	62	55	59	68	65	60	74	66	67
59	71	54	67	64	62	56	71	65	60	60	70	75	66	67	69	55	62
57	74	60	71	59	66	61	66	62	60	62	73	72	64	67	79	60	65
59	69	64	65	62	68	58	62	63	57	61	61	65	68	68	63	62	68
58	69	70	62	68	68	70	67	62	59	61	67	68	70	71	70	61	65
61	65	60	69	61	68	62	68	60	61	62	73	72	60	71	59	56	59
66	66	63	65	66	62	59	67	62	61	63	74	71	67	66	63	61	62
60	70	61	65	61	63	60	65	66	61	62	68	70	59	67	73	56	65
60	62	61	64	60	68	60	68	63	60	57	65	63	57	74	66	62	61
69	71	60	61	63	67	60	64	63	61	57	63	66	65	67	67	63	61
61	69	56	65	59	68	60	61	64	65	63	66	61	71	65	59	62	57
62	73	65	68	59	75	61	69	56	56	64	63	61	69	74	61	55	61
61	71	64	70	63	85	58	63	60	63	71	62	62	67	71	65	63	57
60	67	61	68	56	76	68	65	61	61	63	69	59	71	68	64	60	56
60	68	57	65	59	72	64	69	63	60	59	71	60	69	60	65	57	61
58	72	54	68	60	65	65	71	61	54	63	72	59	64	65	60	65	65
63	65	64	69	60	67	56	68	62	64	62	67	66	68	63	69	62	69
64	65	59	61	65	66	58	66	59	57	61	72	67	63	66	61	52	65

LOT DATA TO ACCOMPANY "WHICH DISTRIBUTION APPLIES?" PAPER FOR 35TH DESIGN OF EXPERIMENTS CONFERENCE (Cont.)

L O T N U M B E R																	
103	104	105	106	107	108	109	110	111	112	113	114	115	116	123	124	125	126
57	67	72	63	75	60	57	56	64	68	59	69	63	63	58	62	67	80
64	57	69	72	58	67	60	59	63	62	65	78	64	70	45	62	68	74
57	63	70	76	63	61	61	62	75	62	61	71	69	70	67	62	69	89
58	76	64	75	66	58	59	66	64	55	62	67	69	70	63	64	65	87
62	64	63	70	67	55	65	72	78	59	61	63	74	64	49	62	58	90
61	69	61	68	67	59	63	58	61	68	68	63	68	72	71	57	65	94
54	77	68	74	73	59	63	59	61	63	60	67	69	65	56	63	67	83
55	74	74	67	76	57	60	62	69	62	59	79	63	70	71	67	65	85
59	63	63	76	64	63	55	72	66	65	64	75	64	59	68	64	64	82
62	61	66	74	61	59	60	63	66	56	66	64	64	57	63	66	63	91
55	64	64	72	64	52	60	64	68	66	65	71	65	57	60	56	64	86
53	65	51	69	63	62	55	57	63	60	59	72	60	62	70	63	69	86
70	66	59	75	63	58	57	61	57	61	65	74	60	66	60	69	66	80
54	69	66	72	63	59	63	55	66	62	66	70	63	58	67	68	69	82
54	73	61	74	67	60	55	65	61	61	62	72	66	53	72	64	70	85
66	56	70	72	66	63	62	66	63	59	60	80	71	62	67	66	65	89
68	63	61	70	74	60	64	60	69	58	58	91	73	38	69	69	67	84
71	66	62	70	62	57	58	60	75	60	58	56	68	63	56	69	66	90
54	71	67	64	62	53	58	60	72	63	63	67	69	61	64	79	65	86
61	79	63	67	63	56	58	60	63	59	60	63	75	65	61	70	65	84

L O T N U M B E R					
127	128	129	130	131	132
60	86	61	64	56	89
59	92	59	73	65	91
56	81	60	81	61	91
61	100	62	76	64	100
59	76	59	73	57	89
59	97	57	67	62	91
65	97	61	65	62	83
58	87	58	83	57	103
59	83	61	69	59	109
65	79	64	70	59	97
66	88	56	74	55	92
67	87	57	81	57	90
68	80	64	67	61	100
66	85	65	74	64	85
64	89	60	71	60	90
67	87	65	87	63	98
71	97	69	78	61	95
59	88	68	77	56	74
65	87	64	88	63	88
62	93	65	70	63	110

STATISTICALLY BASED MATERIAL PROPERTIES

Donald M. Neal and Mark G. Vangel
U.S. Army Materials Technology Laboratory, SLCMT-MRS
Watertown, Massachusetts 02172-0001

ABSTRACT

This paper describes statistical procedures and their importance in obtaining composite material property values in designing structures for aircraft and military combat systems. The property value is such that the strength exceeds this value with a prescribed probability with 95% confidence in the assertion. The survival probabilities are the 99th percentile and 90th percentile for the A and B basis values respectively. The basis values for strain to failure measurements are defined in a similar manner. The B value is the primary concern of this paper.

INTRODUCTION

Many traditional structural materials, which are homogeneous and isotropic, differ from composite materials which have extensive intrinsic statistical variability in many material properties. This variability, particularly important to strength properties, is due not only to inhomogeneity and anisotropy, but also to the basic brittleness of many matrices and most fibers and to the potential for property mismatch between the components. Because of this inherent statistical variability, careful statistical analysis of composite material properties is not only more important but is also more complex than for traditional structures.

This paper addresses this issue by discussing the methodologies and their sequence of applications for obtaining statistical material property values (basis values). A more detailed analysis showing the various operations required for computation of the basis value is presented by the authors in the statistics chapter of the MIL-17 Handbook (ref. 1). The procedures in this handbook required substantial research efforts in order to accommodate various requirements (eg. small samples, batch to batch variability, and tolerance limits) for obtaining the basis values. Guidance in selection of the methodology came from the needs of the military, aircraft industry, and the Federal Aviation Administration (FAA). Some of the procedures include determination of outliers, selection of statistical models, tests for batch to batch variation, single and multi-batch models for basis value computation and nonparametric methods. In figure 1, a flowchart is shown outlining the sequence of operations.

An important application of the basis property value is to the design of composite aircraft structures where a design allowable is developed from this value. The process usually involves a reduction in the basis values in order to represent a specific application of the composite material in a structure (for example, a structure with a bolt hole for a particular test and environmental condition). One common approach in the design process requires the design allowable be divided by the maximum applied stress or strain and the result to be greater than one. The basis value is also used in qualifying new composite material systems to be used in the manufacture of aircraft. In this case, the values are obtained from an extensive test matrix including both loading and environmental conditions. The value also provides guidance in selecting material systems for specific design requirements.

The paper also shows how material strength variability and the number of test specimens can effect the determination of reliability numbers. Methods are presented for obtaining protection against this situation by providing a tolerance limit value on a stress corresponding to a high reliability. A comparison between deterministic and statistical reliability estimates demonstrates the inadequacy of the deterministic approach. A case study is presented describing the recommended procedures outlined in the MIL-17 Handbook for determining statistically based material property values.

RELIABILITY ESTIMATES

Sample Size - Variability

The importance of determining a tolerance limit on a percentile value is graphically displayed in figures 2 and 3. The cumulative distribution function (CDF) of the standard normal (mean equals 0, standard deviation 1) is plotted for sample sizes of 10 and 50, using 25 randomly selected sets of data. In figure 2, for n equals 10, the spread in the percentile is 2.1 for the 10th percentile. In figure 3, for n equals 50, the spread is .7 for the same percentile. The results show the relative uncertainty associated with small sample sizes when computing reliability values. The range in the percentile can also depend on the amount of variability in the data (i.e., the variance).

Often in structural design, a design allowable value is obtained from the basis value. A design allowable is an experimentally determined acceptable stress value for a material (called an allowable stress). The allowable is a function of the material basis value, layup, damage tolerance, open holes, and other factors. It is usually numerically determined for some critical stress region located within the structure. In using the allowable it is required that the critical stress be less than a proportion (margin of safety) of the allowable stress value. Determining a property value from only 10 strength tests using 90% reliability estimates without confidence in the assertion could result in a nonconservative design situation. In order to

prevent this occurrence and provide a guarantee of the reliability value, a tolerance limit (i.e. a lower confidence bound) on the percentile is recommended. The MIL-17 Handbook statistics chapter describes methods for obtaining basis values for a prescribed tolerance limit.

Definition of the B-Basis Value

The B-basis value is a random variable where an observed basis value from a sample (data set) will be less than the 10th percentile of the population with a probability of .95. In figures 4 and 5 a graphical display is shown of the basis value probability density functions for random samples of n equals 10 and 50 respectively. Samples are from the same population as in figures 2 and 3. The vertical dotted lines represent the location of the population 10th percentile ($X_{.10}$). The probability density function of the population is also displayed in the figures. Note that 95% of the time the basis value is less than $X_{.10}$. The graphical display of the basis value density function shows much less dispersion for n equals 50 than for n equals 10; therefore, small sample sizes often result in very conservative estimates of the basis value.

STATISTICAL METHODS - MATERIAL PROPERTY VALUES

Flowchart Guidelines

Since the statistical procedures and the flowchart (figure 1) have been published in the MIL-17 Handbook (ref. 1) and (ref. 2), this paper will only present a brief description of the methods, their purpose, interpretation of results, and the need for following the order of application suggested by the flowchart. The authors have written a computer code which performs the necessary computations for obtaining the basis values as described in the flowchart. The code is available on a diskette, which can be used on various computers including PC's that are IBM compatible. Both the executable and source code are on the diskette. This code is available free of charge from the authors. The flowchart capability was tested by applying the recommended procedures using both real and simulated data sets. The results of the simulations showed at least 95% of computed values were less than the known 10% point, this is consistent with the definitions of 'B'-basis value, see also (refs. 1 and 2).

The flowchart has two directions of operations, one is for the single batch (sample), and the other is for the multi-batch case. A batch could represent specimens made from a manufactured sheet of composite material representing a roll of prepreg material. Published MIL-17 Handbook basis values are usually obtained from five batches of six specimens each.

Initially, let us assume the user of the flowchart has only a single batch or more than one batch but that the batches can be pooled

so that a single sample analysis can be applied. The first operation (see figure 1) is to determine if outliers exist in the data set. A more detailed discussion of outlier detection schemes and applications are published in ref. 3. The method selected is called the Maximum Normed Residual (MNR) procedure (ref. 4) and is published in the MIL-17 Handbook. It is simple to apply and performs reasonably well even though it assumes that the data is from a symmetric distribution. The analysis requires obtaining an ordered array of normed residuals written as

$$NR_i = (x_i - \bar{x})/s, \; i=1,\cdots n \qquad (1)$$

where $\bar{x}$ is the mean, s is the standard deviation (SD), and n is the sample size. If the maximum absolute value of NR_i (MNR) is less than some critical value (CV) (see refs. 1 and 2), then no outliers exist. If MNR is greater than CV, then an outlier X is determined from the largest NR_i value.

Outlying test results are substantially different from the primary data. For example, assume that the data set contains 16 strength values and 15 range from 150 to 200 KSI while the other is 80 KSI. The MNR method would identify the 80 KSI value to be an outlier. The 80 KSI specimen should be examined for problems in fabrication and testing. If a rationale is determined for rejecting this test result, then do not include the outlying test value in the data set when obtaining the basis value. If there is no rationale for rejection, the outlier should remain unless the test engineer believes that a non-detectable error exists.

It is important to identify the existence of outliers but also of equal importance to resist removing the values unless a rationale has been established. Leaving in or arbitrary removal of outlying values can adversely effect the statistical model selection process and consequently the basis value computation. An outlier in a data set will usually result in a larger variance and a possible shift in the mean when compared with the same data without the outlier. The amount of shift and the variance increase depends on the severity of the outlier (distance removed from the primary data set). It is suggested that for small samples (n is less than 20) critical values corresponding to a 10% significance level be used (see refs. 1 and 2) in order to identify outlying values. If the sample is greater than 20, then use the 5% level. It is often difficult to test for outliers when there is a limited amount of data; therefore, the 10% level will provide additional power to detect outliers. This level will also result in more chance of incorrectly identifying outliers. Outliers can be incorrectly identified from data sets with highly skewed distributions; therefore, it is suggested the box-plot method (refs. 1 and 3) be applied for determining outliers in this situation.

Goodness of Fit Test - Distribution Function

Referring to figure 1, the next step is to identify an acceptable model for representing the data. In the order of preference the three candidate models are Weibull, normal, and the nonparametric method. The Weibull model is

$$F_W(x) = 1 - \exp[-(x/\alpha)^{\beta}] \quad , \text{ where} \tag{2}$$

x is greater than 0, α is the scale parameter, and β is the shape parameter, is considered first in the ordering of the test procedures. The Anderson-Darling (AD) goodness-of-fit test statistic (refs. 1 and 5), is suggested for identifying the model because it emphasizes discrepancies in the tail regions between the cumulative distribution function of the data and the cumulative distribution function of the model. This is more desirable than evaluating the distributional assumptions near the mean since reliability estimates are usually measured in the tail regions. The Anderson-Darling test statistic and the observed significance levels computations are described in refs. 1 and 2. Example problems are also shown in ref. 1, demonstrating computational procedures for applying the AD method.

In following the flowchart, if the Weibull model hasn't been accepted as a desired model, then a test for the normal distribution is suggested,

$$F_N(x) = \frac{1}{\sigma(2\pi)^{1/2}} \int_{-\infty}^{x} \exp[-(t-\mu)^2/2\sigma^2]dt \tag{3}$$

where μ is the mean, and σ^2 is the variance. The AD test for the normal model is similar to the test for the Weibull. The procedure used to identify the normal model is also in refs. 1 and 2. It should be noted that for small samples reliable identification of a model to represent the data is difficult unless some prior information of the population is known.

If the Weibull and normal models are rejected, then a nonparametric method can be used to compute the basis value (see flowchart). This method does not assume any parametric distribution as described above. Therefore, model identification is not required, although application of the method can often result in overly conservative estimates for the basis value.

The conventional nonparametric method (ref. 6) requires a minimum of 29 values in order to obtain a 'B'-basis value, and 300 are needed for the 'A'-basis number. This paper presents a method for obtaining 'A' and 'B' basis values for any sample size. The method is a modification of the ref. 7 procedure involving the ordered data values arranged from least to largest with the basis value defined as

$$B = X_{(r)} - K(X_{(r)} - X_{(1)}), \tag{4}$$

where $X_{(r)}$ is r^{th} ordered value and $X_{(1)}$ is the first ordered number. In refs. 1 and 2 tables for r and K values are tabulated for sample sizes n. Note, in the case where 'A' values are required for small sample sizes, it is suggested that nonparametric methods be applied unless some prior information of the model is known. This is because of the limited information available in the lower tail region of the distribution, which can result in erroneous estimates of the reliability numbers. The 'A'-basis value is often used in design where a single load path exists; therefore, it is essential that the value be conservative.

Weibull Method - 'B'-Basis Value

Returning to the sequence of operations as outlined in the flowchart, if the Weibull model is accepted, then determine the basis value from the following relationship

$$B = \hat{\alpha}[\ln(1/P_B)]^{1/\hat{\beta}} \tag{5}$$

where $\hat{\beta}$ and $\hat{\alpha}$ are maximum likelihood estimates of the shape β and scale α of the Weibull distribution. That is, these estimates maximize the likelihood function, which is the product of probability densities (2) evaluated at each of the n data values. Tables for P_B as a function of the sample size n and the code for determining $\hat{\alpha}$ and $\hat{\beta}$ are given in refs. 2 and 3.

Normal Method - 'B'-Basis

If the Weibull model was rejected and the normal model is an acceptable representation of the data, then compute the basis value as

$$B = \bar{X} - K_B S \tag{6}$$

where $\bar{X}$ and S are the mean and SD, and K_B is obtained from tables in refs. 1 and 2.

PROCEDURES FOR MULTIPLE BATCHES

Anderson-Darling Test

If there are more than one batch of data being analyzed, then a significance test is required in order to determine if the batches may be pooled or if a multi-batch statistical analysis is to be applied (see flowchart). Note, the outlier test is to be applied to pooled data prior to testing. The recommended test is the K-Sample Anderson-Darling Test (refs. 1 and 8) which determines if batch to batch variability exists among the K batches. This test is similar to the AD test for identifying acceptable statistical models for representing

data. In the K sample case, paired comparisons are made for the empirical CDF's while the other AD methods compare a parametric CDF with an empirical CDF. In all cases, this comparison involves the integration of the squared difference of the CDF's weighted in the tail region of the distribution. The K-sample AD is basically a two sample test in that each sample (i^{th} batch) is individually compared with the pooled K-1 other batches, repeated K times until each i^{th} batch has been compared. The average of these K two-sample tests determines the K-sample AD test statistic. Tables of critical values and a detailed description of the method and its application is shown in refs. 1, 2, and 8.

If a significant difference is noted among the K batches, then, as shown in the flowchart, a test for equality of variance is suggested using a method in ref. 9. Application of the method, tables, and the necessary relationships for computing the test statistic are given in refs. 1 and 2. The variance test is suggested only as a diagnostic tool. Sample test results that have large variances relative to the other batches may identify possible problems in testing or manufacturing of the specimens. Equality of variance is not required when applying the Modified Lemon method, as discussed below, in the multi-batch case. Although the Modified Lemon method is based on the assumptions of equality of variance and normality, simulation results have shown that these assumptions are not necessary. After testing for equality variance, it is suggested that the basis value be obtained from application of the Modified Lemon method (see figure 1).

The Modified Lemon Method

Composite materials typically exhibit considerable variability in strength from batch to batch. Because of this variability, one should not indiscriminately pool data across batches and apply single batch procedures. The K-sample Anderson-Darling test was introduced into the MIL-17 Handbook in order to prevent the pooling of data in situations where significant variability exists between batches. For the situation where the K-sample Anderson-Darling test indicates that batches should remain distinct, a special basis value procedure has been provided. This method, referred to as the 'ANOVA' or 'Modified Lemon' method, will be discussed next. A detailed description for applying the method is shown in refs. 1 and 2. For a discussion of the underlying theory, see ref. 10, the original Lemon paper, and ref. 11, the Mee and Owen paper which modifies the Lemon method.

The Modified Lemon method considers each strength measurement to be a sum of three parts. The first part is an unknown constant mean. If one were to produce batches endlessly, breaking specimens from each batch, the average of all of these measurements would approach this unknown constant in the limit of infinitely many batches. Imagine, however, that one were to test many specimens from a single batch. The average strength approaches a constant in this situation as well, but this constant will not be the same as for the case where each specimen came from a different batch. The average converges to an overall population mean (a 'grand mean') in the first case, while the

average converges to the population mean for a particular batch in the second case. The difference between the overall population mean and the population mean for a particular batch is the second component of a strength measurement. This difference is a random quantity - it will vary from batch to batch in an unsystematic way. We assume that this random variable has a normal distribution with a mean of zero and some unknown variance which we refer to as the between batch component of variance. Finally, in order to arrive at the value of a particular strength measurement, we must add to the sum of the constant overall mean and a random shift due to the present batch a third component. This is another random component which differs for each specimen in each batch. It represents variability about the batch mean. It also is assumed to have a normal distribution with a mean of zero and an unknown variance, which is referred to as the 'within batch' component of variance.

The 'Modified Lemon' method uses the data from several batches to determine a material basis property value which provides 95% confidence on the appropriate percentile of a randomly chosen observation from a randomly chosen future batch. This basis property provides protection against the possibility of batch-to-batch variability resulting in future batches which have lower mean strength than those batches for which data are available.

To see what this means, imagine that several batches have been tested and that this statistical procedure has been applied to provide a 'B'-basis value. Now, imagine that you were to get another batch and test a specimen from it. After this you obtained still another batch and tested a specimen from it. If you were to repeat this process for infinitely many future batches, you would obtain a distribution of strength measurements corresponding to a randomly chosen measurement from a random batch. You can be 95% certain that the basis value which you calculated originally is less than the tenth percentile of this hypothetical population of future measurements. This is the primary reason why the Modified Lemon method is advocated by the MIL-17 Handbook - it provides protection against variability between batches which will be made in the future through the use of data which is presently available.

An illustrative example of this method applied to nine batches of material is shown below. The data sets did not pass the K-sample AD test for pooling. Let the batches be

1	2	3	4	5	6	7	8	9
61.3	66.5	66.0	61.9	68.9	75.8	72.8	71.9	68.7
68.5	64.7	72.7	68.0	65.0	75.2	75.0	71.0	76.3
62.5	64.9	67.1	63.3	70.9	71.5	66.3	69.5	76.6
66.0	65.2	67.7	74.6	65.4	69.6	69.5	69.5	66.2
66.6	70.3	65.7	66.2	66.5	66.1	71.9	72.6	72.4
64.8			68.2	64.9			74.6	72.8
69.5			69.1					109.6

with a single outlier, 109.6 determined from MNR method. Let's assume 109.6 was an incorrect test result and replaced by 69.6, a corrected test value.

After a substantial amount of computation (see refs. 1 and 2) involving sums of squares, within batch and between batch variances, non-central t distribution, etc., the 'B'-basis value is

'B' = 60.93

The summary statistics are

Batch	n_i	$\overline{X}_i$	S_i
1	7	65.60	2.99
2	5	66.32	2.33
3	5	67.84	2.84
4	7	67.33	4.17
5	6	66.93	2.45
6	5	71.64	4.03
7	5	71.10	3.33
8	6	71.52	1.96
9	7	71.80	3.88

It should be noted the value of 60.93 is lower than 61.9 of nonparametric solution from the pooled sample. The Modified Lemon method can be overly conservative (low basis values) in order to guarantee 90% reliability with 95% confidence. The number of batches and the variability between and within the batches effect the computation of the basis value. If there are few batches and large between batch variability with small within batch variability, then this situation could result in very low basis numbers depending on the amount of variability and number of batches.

In figure 6 results from application of flowchart procedures are shown for three batches of five specimens of AS4/Epoxy material tested in compression. In this case, the mean strength values show a small amount of variability while there is a relatively large spread within each data set. 'B'-basis results from the flowchart application are for the following: ANOVA (Modified Lemon), Weibull, Normal, Lognormal, and nonparametric methods. Not included in the flowchart results are a list of assumptions that were violated. The results show a small difference in basis values except for the nonparametric solution which has the low value of 167.1. The Weibull method was suggested since it passed the K-sample AD test and the AD goodness-of-fit test. The relatively large within batch variances and small differences in mean values made it possible to pool the batches.

Figure 7 shows another result of computing the 'B'-basis values using the ANOVA, Weibull, and normal methods applied to another three selected batches from same population as in figure 6. The ANOVA result of 15.7 KSI is substantially lower than those from the other

two methods. Unfortunately, this is a result of a large difference in mean values preventing pooling of the batches resulting in the required ANOVA application. The large difference in mean values in addition to relatively small within batch variability resulted in this extremely low basis value. A 'B' value of 6.5 was obtained from the simple normal analysis using the three mean values. The result shows that for this example the ANOVA method primarily depends on the batch means. The above results would suggest obtaining more batches or investigating testing and processing procedures.

In figure 8, results are shown for the case of randomly selecting another batch from the same population described in figure 7. In this case the ANOVA result shows a value of 105.4 KSI which is substantially larger than the 15.7 KSI recorded for the three batches. The importance in having a larger number of batches is shown from these results in figures 7 and 8. Also, with more data available, the pooled results for Weibull and Normal model also resulted in less conservative values.

Figure 9 presents results showing where a substantial amount of within batch data is not necessary. In case 1, the ANOVA results for three batches of 100 data values each, resulted in 154.9 KSI while for case 2, three batches of ten each, a 'B'-basis value of 152 KSI was obtained. This result emphasizes the importance of being able to obtain more batches rather than increasing the batch size. However, the ANOVA results in figure 6 show three batches can provide reasonable results similar to pooled results if small differences in mean values relative to batch variances exist. Note that for very large batch sizes, the K-sample AD test can reject pooling of data even though there is a small difference in mean values. This rejection is statistically correct, but the user of the flowchart may consider the difference in the batch means not of engineering importance. In this case the user can make the decision of pooling or not pooling, since there will be a small difference in basis values from pooled or unpooled results. If there are large batch differences and the ANOVA method is suggested from the flowchart, then adding more batches can reduce the conservatism. The ANOVA method is a random effects model which determines a basis value representing all future values obtained from the same material system and type of test. In order to provide this guarantee in the presence of large batch to batch variability, there is the potential for it to be overly conservative which was shown in figure 7.

Reliability at Basis Stress Value

Figure 10 conceptually describes the statistical reliability of a simple structure in tension as it relates to the 'B'-basis applied stress value. In the example shown in the figure, ten percent of all the specimens (structures) will fail when subjected to load S. This statement should be incorrect at most one time in twenty (95% confidence). S is the 'B'-basis value obtained from strength (failure load) measurements from specimens of similar material and geometry. This statistical guarantee that at most 10% of the specimens will fail

can provide the engineer with a quantitative number for selecting and applying material in composite material structures. This is unlike the conventional deterministic property value approach which is an ad hoc procedure that reduces the mean strength measurements in order to obtain some design value which can result in a potentially over or under design situation. In applying the statistical basis value, it is assumed the material, geometry, and loading conditions in the structural design situation is similar to those obtained from the strength measurements. This is also true for deterministic property value applications. In the following sections the inadequacies of the deterministic approach are discussed in more detail.

Reliability Values Statistical vs. Deterministic

In figure 11 the results of a simulation process involving the random selection of ten values from population of 191 strength measurements repeated 2,500 times are graphically displayed. For each simulation a design number or material property value is obtained from each of the three procedures $\bar{X}/2$, $(2/3)\bar{X}$, and the MIL-17 flowchart. The mean value of the data set is $\bar{X}$. The reliability values, as shown in the figure, are obtained by evaluating the population probability distribution fit to the 191 values at the design numbers.

In the case where the mean is reduced by a factor of 1/2, the strength values are very low (90 KSI), and the reliability is extremely high (1.0). The engineer may not be able to afford such a high reliability value of 1.0 (to twenty significant digits) at the expense of having design values as low as 90 KSI when mean strength is 180 KSI. The factor of 2/3 increases the design value but reduces the reliability to approximately .999. The flowchart 'B'-basis calculation provides higher strength values with acceptable reliability numbers. The other two procedures show an element of uncertainty by depending on the chosen factor. If the engineer used the factor of 1/2, this would result in an extremely over design situation require either rejection of the material or the design. Alternatively, if the engineer used the mean strength as design number, the reliability would be reduced to .5, although strength values would be much higher. The flowchart procedure removes the uncertainty by providing a guaranteed minimum reliability of .90 without unnecessarily reducing the basis value. The minimum reliability can be increased to .99 if necessary by using 'A'-basis computations as outlined in the MIL-17 Handbook.

Effect of Variance on Reliability Estimates

In figure 12 the effects of variance differences as they relate to reliability estimates are shown from a simulation process. This involved randomly selecting ten values from each of two separate normal distributions with same mean of 100 and different SD's of 5 and 25 repeated 2,500 times. The reliability values are obtained in a similar manner as described in the previous section, except the probability values were obtained from the normal distribution. In the case

where the SD is 5, there is very little dispersion in the reliability values. Again, the design number from $\bar{X}/2$ is substantially lower than the basis value using the flowchart process, although the reliability is very high for this number. In comparing this with the results using SD of 25, a substantial increase dispersion of the reliability values particularly for the basis results using flowchart methods. The flowchart results show similar reliability estimates for both SD's of 5 and 25, although for the $\bar{X}/2$ the reliability has been reduced substantially from twelve nines to .96. This is the result of the deterministic ($\bar{X}/2$) approach being independent of variance. This is not an issue if 50% reliability is required, but for 90% reliability, variability is important. Dividing the mean by two can be nonconservative for situation when the distribution has a large spread (long tail). In order to make adjustment for this situation, the flowchart method (basis value) is suggested. See results in the figure where the basis value adjusts to a lower level but maintains the same range for the reliability estimates. The basis value will guarantee a reliability by adjusting the design value while the safety factor approach cannot guarantee reliability. This result suggests using the basis method if it is important to maintain a certain level of reliability. The overall issue is that the flowchart methods will provide property values with specified reliability with 95% confidence while the deterministic approach is an ad hoc approach with no control of the resulting reliability estimates.

CONCLUSIONS

This paper is an exposition of the statistical procedures described in the MIL-17 Handbook for obtaining material property values. Its primary goal was to introduce the MIL-17 statistics chapter to the users so that they may use it more effectively. The methods and the sequence of operations suggested by the statistics chapter flowchart were analyzed with respect to their effectiveness, purpose, and limitations. By following the flowchart procedures, guidance is provided to the user so that reasonably accurate property values may be obtained without relying on ad hoc schemes which could potentially result in either excessively low or high values.

Each method and its order of application were discussed with respect to their specific purpose, such as model identification, batch to batch variability recognition, outlier detection, and the basis value computation. There are situations where low basis values will result, not because of limitations in the statistical procedures but are usually the result of very large or small data sets, large batch to batch variations, or model recognition.

The comparison between the statistical reliability and the deterministic approach showed a preference for statistics since it was able to guarantee a specified reliability in contrast to a deterministic method which is primarily an ad hoc process resulting in considerable uncertainty as to the corresponding reliability estimates. Finally, the authors have attempted to provide a satisfactory definition of a

statistically based material property value by introducing the tolerance limit concept and its importance. A number of illustrations were presented showing the advantage of the tolerance limit over the deterministic approach.

ACKNOWLEDGEMENT

The authors wish to thank Lucy Ohannesian of MTL for preparing this manuscript and to Joseph Soderquist of the FAA for his guidance in the selection of the statistical problems described in this paper. The assistance of Prof. Bernard Harris of the University of Wisconsin in developing the statistical methodology is also acknowledged.

REFERENCES

1. MIL-HDBK-17B, Polymer Matrix Composites, vol. 1, February 1988. Naval Publications and Forms Center, 5801 Tabor Avenue, Philadelphia, Pennsylvania 19120.

2. Neal, D. M.; Vangel, M. G.; and Todt, F.: Determination of Statistically Based Composite Material Properties. In Engineered Materials Handbook, vol. 1, 1987, Composites, ed. Cyril A. Dostal, American Society of Metals Press, Metals Park, Ohio.

3. Neal, D. M. and Spiridigliozzi, L.: An Efficient Method for Determining "A" and "B" Allowables. ARO 83-2, in Proceedings of the 28th Conference on the Design of Experiments in Army Research, Army Research Office, 1983, pp. 199-235.

4. Stefansky, W.: Rejecting Outliers in Factorial Designs. Technometrics, vol. 14, 1972, pp. 469-479.

5. Anderson, T. W. and Darling, D. A.: A Test of Goodness-of-Fit. J. Am. Statis. Assoc., vol. 49, 1954, pp. 765-769.

6. Conover, W. J.: Practical Nonparametric Statistics. John Wiley and Sons, 1980, p. 111.

7. Hanson, D. L. and Koopmans, L. H.: Tolerance Limits for the Class of Distributions With Increasing Hazard Rates. Annals of Mathematical Statistics, vol. 35, 1964, pp. 1561-1570.

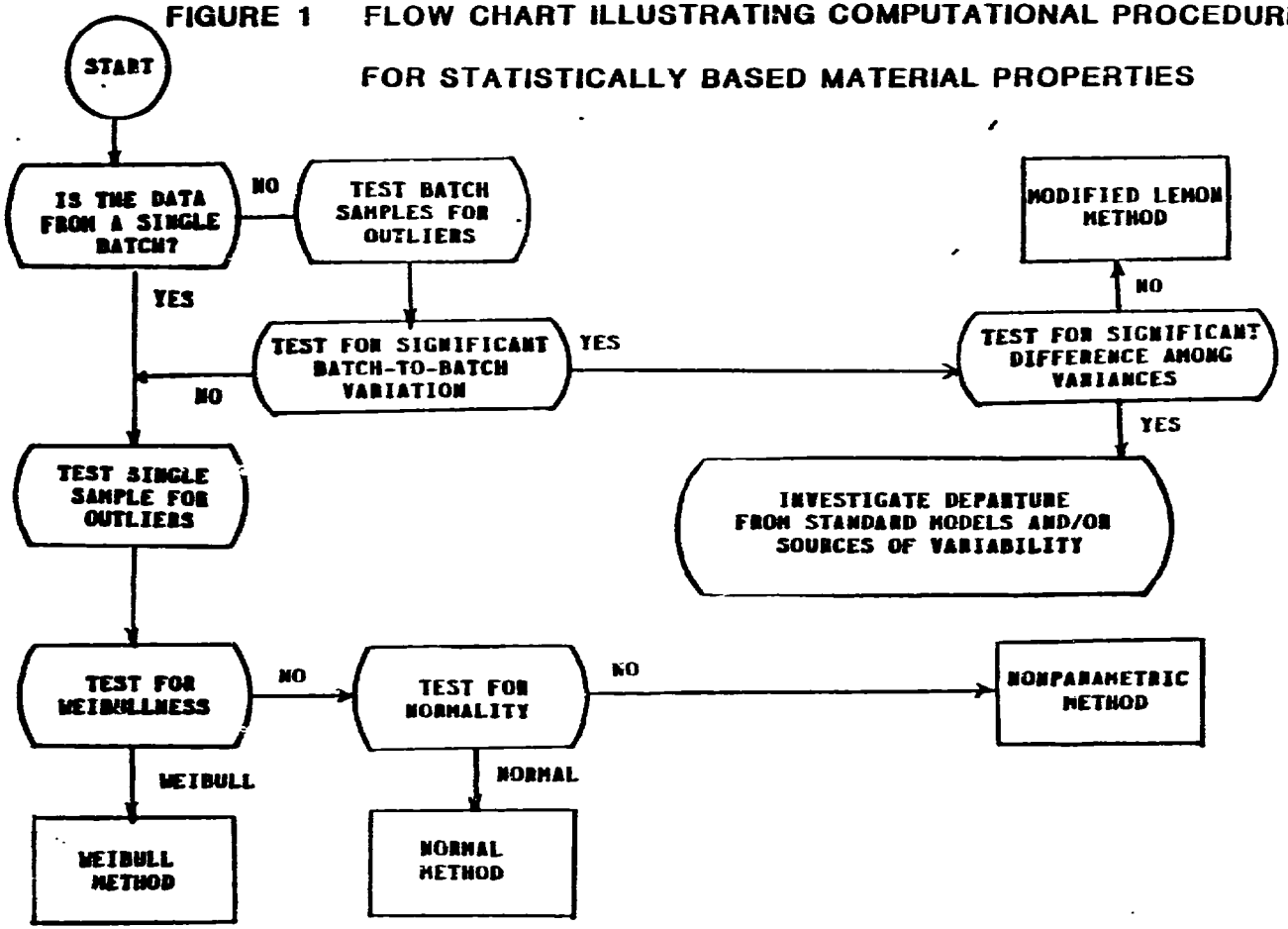
8. Scholz, F. W. and Stephens, M. A.: K-Sample Anderson-Darling Tests. J. Am. Statis. Assoc., vol. 82, 1987, p. 918.

9. Lehmann, E. L.: Testing Statistical Hypotheses. John Wiley and Sons, 1959, pp. 274-275.

10. Lemon, G. H.: Factors and One-Sided Tolerance Bounds for Balanced One-Way ANOVA Random Effects Model. J. Am. Statis. Assoc., vol. 72, 1977, pp. 676-680.

11. Mee, R. W. and Owen, D. B.: Improved Factors for One-Sided Tolerance Limits for Balanced One-Way ANOVA Random Effects Model. J. Am. Statis. Assoc., vol. 78, 1983, pp. 901-905.

FIGURE 1 FLOW CHART ILLUSTRATING COMPUTATIONAL PROCEDURES FOR STATISTICALLY BASED MATERIAL PROPERTIES



FIGURES 2 AND 3 SAMPLE SIZE EFFECT ON RELIABILITY

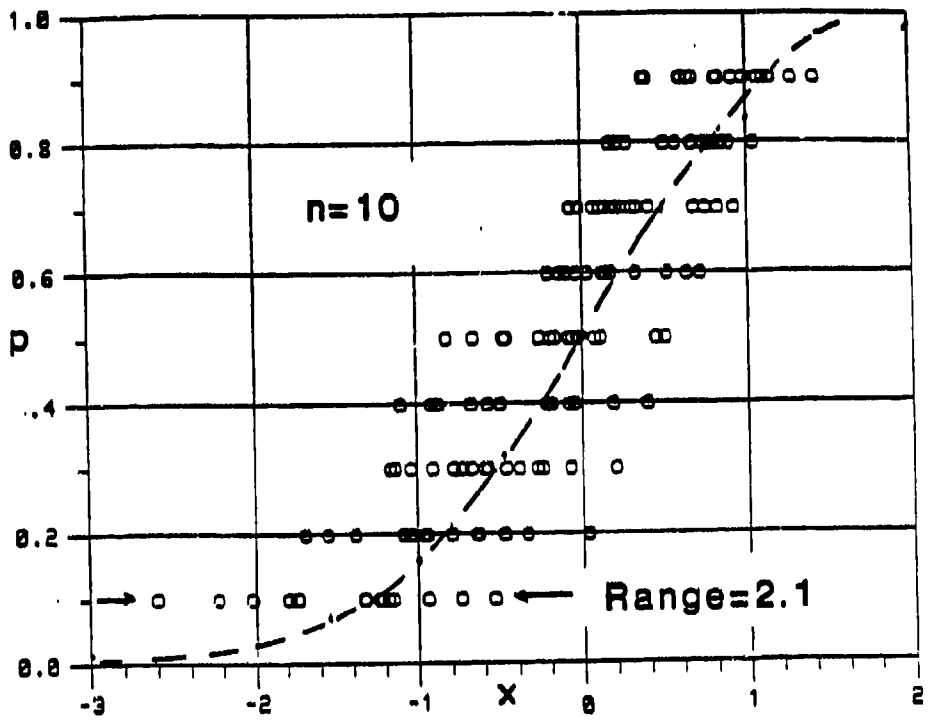
Random data sets of size n from a Normal distribution.

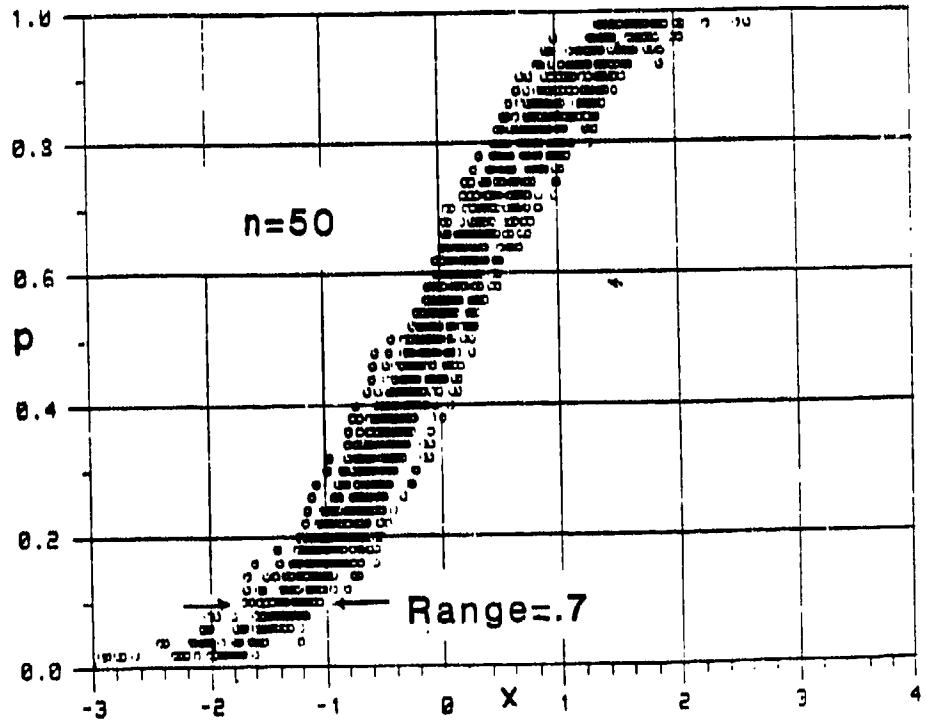
Example:
Empirical CDF

I	x	I/n (=p)
1	-2	.10
2	-1	.20
3	0	.30
4	1	.40
5	1.5	.50
6	1.8	.60
7	1.9	.70
8	2.1	.80
9	2.3	.90

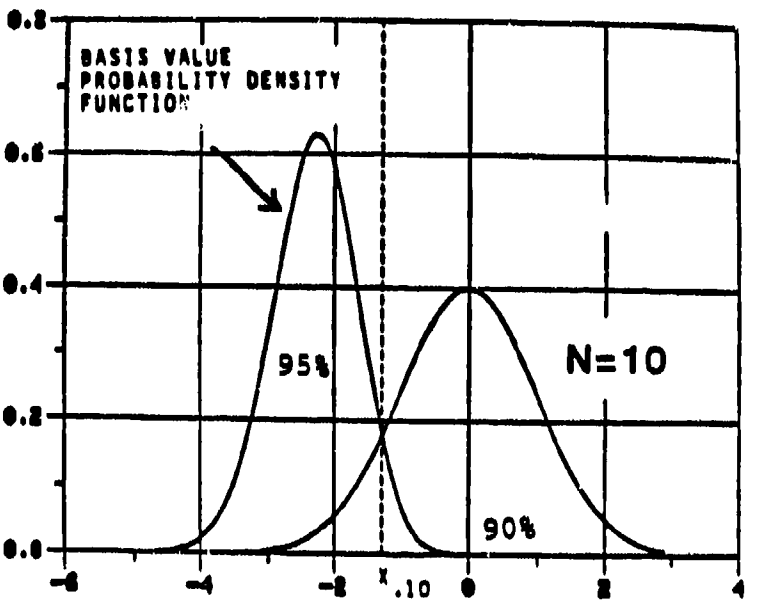
mean=0

sd=1





FIGURES 4 AND 5 BASIS VALUE PROB. DENSITY FUNC.



Density

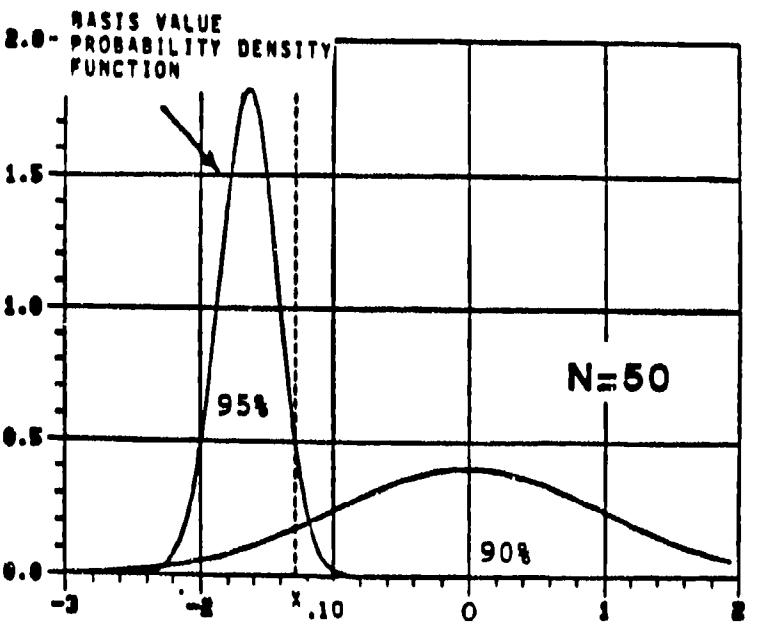
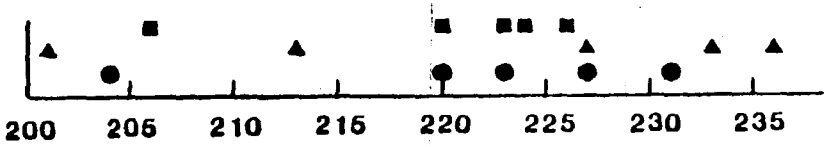


FIGURE 6 EXAMPLE OF BASIS VALUE CALCULATION
NEGLIGABLE BATCH-TO-BATCH VARIABILITY

Batch 1: Mean strength = 221 Ksi
Batch 2: Mean strength = 222 Ksi
Batch 3: Mean strength = 220 Ksi



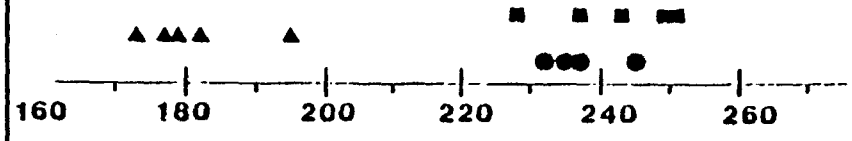
METHOD	BASIS VALUE	
ANOVA	202.6	Ksi
Weibull	*196.5*	*Ksi*
Normal	199.1	Ksi
Lognomal	199.6	Ksi
Nonparametric	167.1	Ksi

The Weibull result is recommended by the Flowchart.

(3 batches of 5, AS4/Epoxy compression)

FIGURE 7 EXAMPLE OF BASIS VALUE CALCULATION
SUBSTANTIAL BATCH-TO-BATCH VARIABILITY

Batch 1: Mean strength = 181 Ksi
Batch 2: Mean strength = 236 Ksi
Batch 3: Mean strength = 241 Ksi



METHOD	BASIS VALUE	
ANOVA	*15.7*	Ksi
Weibull	161.9	Ksi
Normal	159.3	Ksi

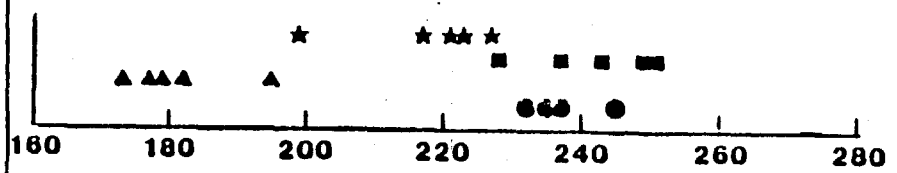
The ANOVA result is recommended by the Flowchart. Normal analysis using only the three mean values gives a B-basis value of 6.5. Either reject the material as too variable or obtain more batches.

(3 batches of 5, AS4/Epoxy compression)

FIGURE 8 EXAMPLE OF BASIS VALUE CALCULATION

THE EFFECT OF AN ADDITIONAL BATCH

Batch 1: Mean strength = 181 Ksi
Batch 2: Mean strength = 236 Ksi
Batch 3: Mean strength = 241 Ksi
Batch 4: Mean strength = 217 Ksi



METHOD	BASIS VALUE
ANOVA	*105.4* Ksi
Weibull	170.3 Ksi
Normal	170.0 Ksi

The ANOVA result is recommended by the Flowchart. A single additional batch increased the basis value from 15.7 Ksi to 105.4 Ksi.

(4 batches of 5, AS4/Epoxy compression)

FIGURE 9 THE EFFECT OF INCREASED BATCH SIZE:

SUBSTANTIAL BETWEEN-BATCH VARIABILITY

Case 1:
T300/Epoxy Unidirectional Tension
3 batches of 100 specimens each

Method	Basis Value
ANOVA	*154.9*
Weibull	171.7
Normal	175.7

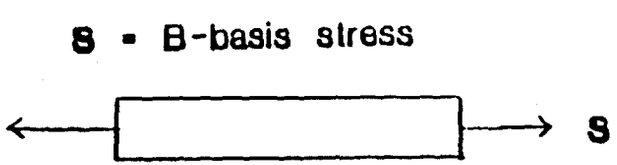
Case 2:
One random dataset of 10 from each of the above three batches.

Method	Basis Value
ANOVA	*152.0*
Weibull	165.7
Normal	172.5

The ANOVA method is recommended by the Flowchart. Note that there is little difference between basis values for batch sizes of 10 and basis values for batch sizes of 100.

FIGURE 10 RELIABILITY AT BASIS STRESS: STATISTICAL VS. DETERMINISTIC

The reliability of a test specimen at the B-basis stress should be high. For a statistically based B- basis value calculated from a procedure appropriate to the data, this reliability is guaranteed to be at least 90% (i.e., 90% with 95% *confidence*).



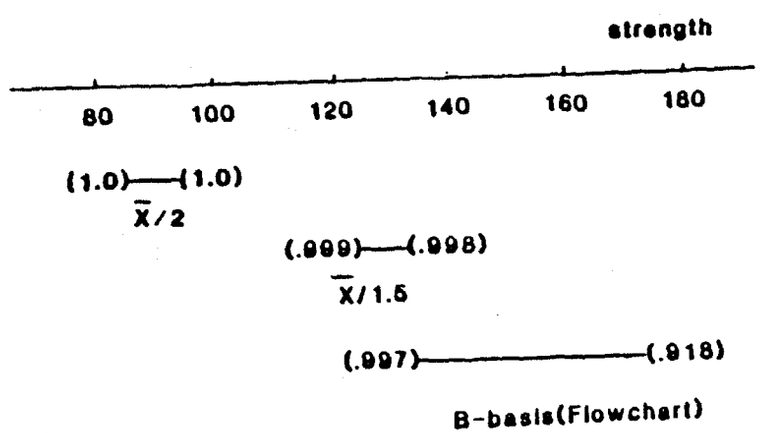
N specimens of which F fail at or below stress S.

Estimated reliability at B-basis stress (N-F)/N

FIGURE 11 RELIABILITY / STRENGTH COMPARISON: A CASE STUDY - STAT. VS. DETERMINISTIC

Population: 191 strength values

Dataset: 10 specimens chosen 2500 times randomly

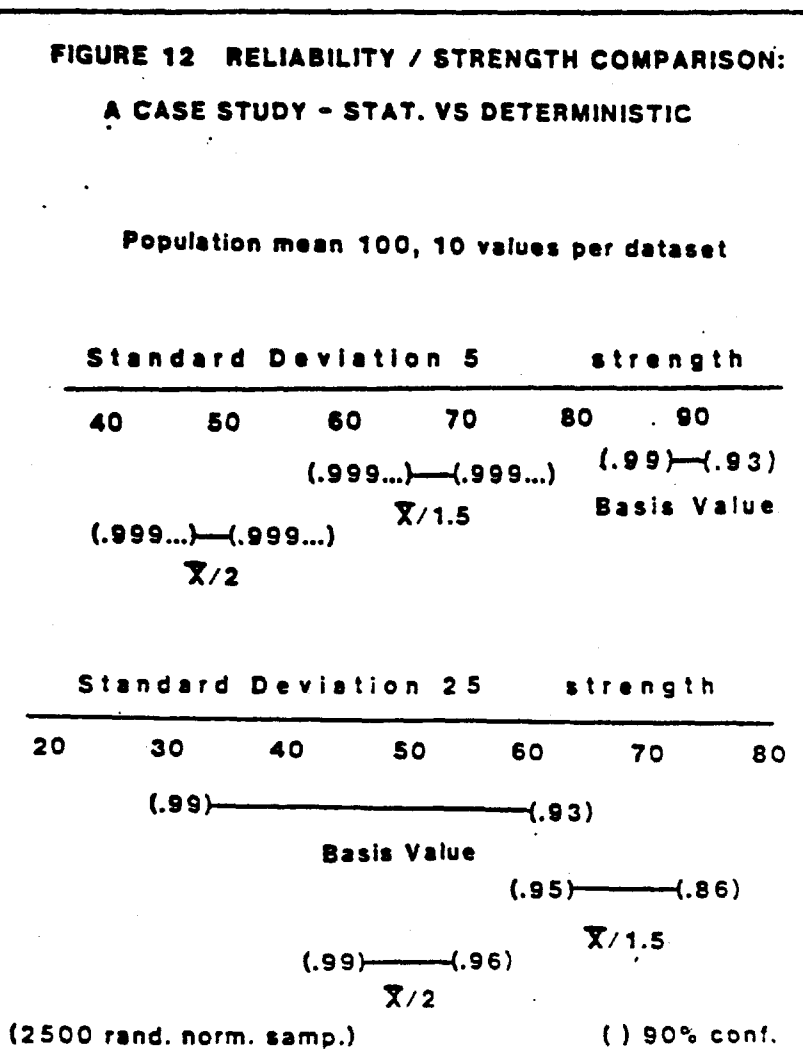


() Reliability Values

(T300/Epoxy Unidirectional) TENSION

FIGURE 12 RELIABILITY / STRENGTH COMPARISON:

A CASE STUDY - STAT. VS DETERMINISTIC



STATISTICAL CULTURE: PROMOTING THE PRACTICE OF STATISTICS

Emanuel Parzen

Department of Statistics, Texas A&M University

Abstract

This paper proposes a framework, called Statistical Culture, for studying the practice of statistics with the aim of improving the health of statistical science as measured by how well citizens and scientists use it as a tool in their daily life and research. We identify a paradigm for lifelong learning based on identifying five (parallel, non-hierarchial) levels of statistical literacy: consumer, applier, consultant, collaborator, theorist. We support accreditation of statistical literacy. We make recommendations for how statisticians can promote public recognition of the importance of statistics, statistical literacy, and interaction between researchers and statisticians. We propose "solutions" to the use of statistics as a scientific method by research which aims to unify and guide thinking about the diversity of statistical methods and theories.

Contents: Statistical Culture as a Paradigm for Lifelong Learning, Solutions, Problems, Levels, Excellence, Statistical Culture Levels Theorem, Olkin-Sacks Report, Statistical Culture Applications Theorem, Statistical Culture Research Problems.

KEYWORDS: Foundations, Teaching, Statistical Literacy, Statistical Science, Unification of Statistical Methods, Statistical Culture.

Research supported by U. S. Army Research Office.

October 1989

STATISTICAL CULTURE: PROMOTING THE PRACTICE OF STATISTICS

Emanuel Parzen

Department of Statistics, Texas A&M University

STATISTICAL CULTURE AS A PARADIGM FOR LIFE LONG LEARNING: The health of a society is becoming increasingly dependent on its statistical literacy, and how statistics is practiced. Modern society is data-rich and has an ever-increasing need to understand how data becomes information (useable knowledge). The goal of continuous improvements of quality of processes involved in the delivery of products or services requires that decisions be based on the information in data, not just on opinions or guesses; this is the main recommendation of the philosophy of Ed Deming (see Mann (1988), p. 15).

This paper proposes that the practice of statistics at any of its levels should be a lifelong endeavor characterized by the features that are being advocated as the requirements of paradigms for lifelong learning that will be required in the 21st century (according to John Sculley (1989), p. 1057):

- "It should require rigorous mastery of subject matter under expert guidance.
- It should hone the conceptual skills that wrest meaning from data.
- It should promote a healthy skepticism that tests reality against multiple points of view.
- It should nourish individual creativity and encourage exploration.
- It should support collaboration.
- It should reward clear communciation.
- It should provoke a journey of discovery.
- And above all it should be energized by the opportunity to contribute to the total of what we know and what we can do."

The study of how to achieve the lifelong learning process required for the practice of

statistics is called "statistical culture."

This paper seeks to show the important role of "statistical culture" in the practice of statistics. It supports the concept of accreditation of statistical literacy at various levels.

The challenge for statistical education will be to find ways of bringing to the process of instruction the passion for discovery that drives excellent statistical thinking.

SOLUTIONS: Statistical culture (the study of the practice of statistics) has goals of elegance and utility. The elegance of statistical culture is obvious; it enhances the fun of doing statistics.The utility of the study of the culture of statistics is to motivate statistical "steersmanship", developing consensus about (and implementing) the actions needed for continuously evaluating and improving the health of the discipline and profession of statistics.

Statistical culture can be said to be the study of the maps (geography, current history) of statistics, rather than its ancient history (as in the history of statistics up to 1900). It is the study of the maps of statistics from the point of view of understanding its current state of the art and influencing its future development.

Statistical culture can be defined to be the study of:

how statistics is, and ought to be, practiced;

where statistics has applications (see Table 1) and who is doing the applying;

what to teach in statistics courses;

why statistics works;

when are competing probability models and statistical methods successful;

accreditation of statistical literacy (rather than competency) at various levels.

To promote the practice of statistics, statistical culture seeks:

1. To develop maps of statistical methods which will help applied statisticians to strive for continuous improvement of methods, to learn new methods to consider as alternatives, to compare competing methods, to more confidently obtain conclusions from

comparisons of the results of competing methods of statistical data analysis of data of a certain type, to obtain problem-driven results from methods-driven results, to obtain substantive conclusions from data for which prior substantive knowledge was not available.

2. To develop maps of statistical theories which help theoretical statisticians to define frontiers of research and thus understand the sense and purpose of research which otherwise may seem unfocused and unmotivated.

3. To develop maps of the relations between statistics and other fields of knowledge and research which will help interactions between statisticians and researchers in other disciplines provide more recognition to the research contributions of statisticians.

4. To develop maps of the contributions that statistical literacy and the practice of statistics can make to a nation's quality of life and world competitiveness.

5. To organize (each year, in each community) Statistical Science Awareness Days to promote the practice of statistics and public recognition of outstanding statisticians.

Statistical culture (which develops unifications, maps, frameworks) is urgently needed in order to improve the image of statistics among scientists and professionals. It would provide the ability to objectively recognize by suitable awards more statisticians as "outstanding" contributors to the missions of their organizations as well as to the discipline and the profession of statistics.

Unification of methods is one of the important facets of the use of the scientific method in any field of research (and therefore, a fortiori, in statistics). Unification of statistical methods does not prevent statisticians from using ad hoc solutions (which many claim is their preferred approach) but rather encourages and guides such methods by clarifying the methods available which may be chosen ad hoc; therefore the ultimate goal of research (such as Parzen (1989)) on Grand Unified Theories of Statistical Methods, denoted GUTS, is "grand unified ad hockery".

PROBLEMS: Statisticians are increasingly aware that there are urgent problems in the discipline and profession of statistics; we argue that problems can be solved if they are discussed using scientific methods and a framework for the "culture" of statistical practice. Examples of such problems are: declining enrollment of statistics doctoral students, difficulty of attracting young people into a career in statistics, teaching statistics to engineers (Penzias (1989)), misunderstanding of the role of statisticians in quality control and quality manufacturing (Hahn (1989)), expressions of dissatisfaction in the profession of statistics about the appreciation and utilization of statisticians (Boroto and Zahn (1989) and McPherson (1989)), failure of leading statisticians to continuously promote statistical culture (to be providing leadership to the study of promoting the practice of statistics), failure of many statisticians to be literate at appropriate levels in a diversity of statistical methods (including time series analysis).

LEVELS: We believe that one can apply the scientific method to the study of statistical culture (the investigation of how statistics is, and ought to be, practiced); answers to such questions should not be based on prejudices but on a consensus of the philosophical writings of successful statisticians. From recent literature about statistics (Bodmer (1985), McPherson (1989)) one can conclude the following first step in drawing a map of the practice of statistics (which we state below in more detail as the Statistical Culture Levels Theorem).

The practice of statistics occurs at three levels of understanding and practice:

popular,

science-related professionals, and

professional statisticians;

further the practice of statistics by statisticians can be divided into three levels:

consulting

collaboration

theory and methods.

EXCELLENCE: Statistical culture aims to provide a framework which stimulates statisticians to understand and applaud each other's work (indeed, there seems to be too much joy in "statistician bashing"); this may be a general failing of human nature but it seems to be an urgent problem for statistics. The use of the word "level" should not be interpreted as implying a vertical or series structure, with activity in statistical theory at the top. The levels form a horizontal or parallel structure; it cannot be emphasized enough that the understanding required in each level involves different aspects of the practice and methods of statistics. A possible analogy is the saying: "Use the talents you possess; for the woods would be very silent if no birds sang except the best."

Statistical culture does aim to support the search for excellence. Criteria should be developed to rate good statistical practice as either average, superior, or exceptional; one criterion is whether it is done at the level of "what," "how," or "why".

STATISTICAL CULTURE LEVELS THEOREM: CONSUMER, APPLIER, CONSULTANT, COLLABORATION, THEORY AND METHODS DEVELOPMENT. To promote the practice of statistics, we propose that it is useful to identify five levels of practice, defined as follows.

I. Statistical consumer:

knows definitions of statistics;

appreciates the concept of variability (distribution of outcomes);

has the ability to understand statistical models and graphical presentations of data analysis;

does not have a working knowledge of statistical methods or the ability to carry out a statistical analysis;

appreciates the role of statisticians in the battle for statistical literacy (competence in understanding, applying and advancing statistical reasoning).

Statistical literacy at the consumer level can be defined to be knowing that public

policy should be based on answers to the questions: "What can happen? What are the odds (probabilities)? How do you know the odds?"

II. Statistical applier. Distinguish two levels:

II(A). knows basic statistical methods used to determine and obtain needed information; ability to use menu driven statistical computing packages; fits all problems into convenient routine statistical conceptualizations;

II(B). ability to use command driven statistical computing environments; understands the assumptions underlying statistical methods and can adapt statistical methods to provide ad hoc methods for problems at hand;

Scientists and engineers involved in research or development should be statistical appliers; those that become more statistically self-sufficient can become more responsible to be their own statistical consultants.

III. Statistical consultant:

skilled in transforming data into information;

has the ability to examine facts and serve as referees of statistical analyses;

aware of the most modern statistical methods;

not actively involved in the scientific language and perspective of the problems being studied so that conversation between client or customer and consulting statistician is less a dialogue and more a monologue;

requires abilities to interview clients to obtain an understanding of their problems, and to communicate with clients by oral presentations and written reports;

often advised to use simple techniques for scientists unable to appreciate subtleties of statistics;

helps contribute to research on the consulting process.

IV. Statistical collaborator:

statistician is a collaborator on the project and is a catalyst and potential advocate of

actions and directions to be pursued in the project;

collaborative research often (if not always) leads to joint publications and/or joint research grants;

has mathematical training adequate to understand the philosophy and vigor of statistical methods but not completely the rigorous proof of their theory;

has ethical, administrative, and diplomatic skills, especially those required for large scale and long term research projects;

helps contribute to research on the collaboration process.

V. Statistical theorist:

inevitably mathematically well trained,

seeks to develop and teach the logical structure of statistical methods, to understand how they are born and how they die, how they can be made to work better and why they work;

basic research in general methods that provide analogies between applications;

fundamental research in analogies between methods (patterns which general methods share with other general methods);

mathematical research on the properties of statistical methods can be considered another level within the theory level.

OLKIN-SACKS 1988 REPORT: The distinction between consulting and collaboration is based on how "equal" the statistician is regarded as a member of the research team. Olkin and Sacks (1988) used the names "advisory collaboration" and "interactive collaboration" (or Type A and Type B) for what we call "consulting" and "collaboration". We quote the report (p. 12):

"Typically, the statistician engaged in advisory work will adapt existing methodology to the problem at hand and create computable versions of known techniques. Another mode of collaboration is much more interactive in nature and involves work to develop

novel techniques and methods to deal with broader substantive questions. This second type of collaboration leads to research on statistical issues that may subsequently advance knowledge both in the substantive field and in statistics itself.

"The survey responses indicated a high frequency of Type A research, while sounding a common theme that Type B research does not receive sufficient time, money, or recognition of its value. The short-run 'advisory consultation' rarely becomes the 'long-range interactive collaboration.' Yet it is the interactive mode that has the greater potential to break new ground and lead to statistical innovations of far-reaching significance for the future conduct of science, and it is this type of collaboration that the panel feels must receive the attention of the disciplines and of NSF and other funding agencies."

STATISTICAL CULTURE APPLICATIONS THEOREM: Another map required to guide the practice of statistics, called a Statistical Culture Applications Theorem, is given in Table 1 which lists disciplines represented in cross-disciplinary research involving collaboration by faculty members in "statistics programs" in universities. The fields and percentages are vaguely adapted from Table 5 of the Olkin-Sacks report. The conjectured percentages are intended to motivate passsionate discussions (and, eventually, research). An interesting research program is to investigate the proportion of new degrees in statistics that take employment to apply statistics in each discipline listed in Table 1.

The interests of statisticians may also be studied by investigating the distribution of 1987 doctorates among broad fields of statistics (see Cox, Voytuk, and Hart (1989)):

Probability and Math Stat	143
Biometrics and Biostatistics	37
Psychometrics	9
Econometrics	25
Social Sciences Statistics	49
TOTAL	263

The racial/ethnic composition of mathematical doctorate degree recipients in the period 1975 to 1986 was as follows:

	White	Black	Hispanic	Asian
Math Sciences, total	89.8%	1.4%	1.4%	7.1%
Prob & Math Stat	85.9%	1.5%	1.5%	10.9%

The percentage of degrees to foreign citizens is 40% in statistics and 45% in mathematics. The percentage of math-science doctorates working in education is 50% for statistics and 60% for mathematics; 25% of statistics doctorates are university faculty members.

Table 1: Disciplines Where Statistics is Applied
Disciplines Represented in Statistical (and Time Series Analysis)
Cross-Disciplinary Collaborative Research

(Conjectured Percentage of Statisticians in Universities Involved in Collaboration)

Health and Life Sciences (25%, 25%)
- Medicine
- Public Health and Epidemiology, Biostatistics
- Biology
- Ecology
- Fisheries and Wildlife
- Environmental Sciences
- Pharmacology and Toxiocology
- Genetics
- Entomology
- Forest Science
- Physiology

Engineering and Mathematical Sciences (15%)
- Engineering
- Computer Sciences
- Operations Research and Reliability
- Mathematics
- Signal Processing
- Image Analysis and Pattern Recognition
- Industrial Statistics
- Defense Statistical Standards
- Hydrology

Behavioral and Social Sciences (15%)
- Psychology, Cognitive Sciences
- Economics, Econometrics
- Education
- Sociology
- Political Science
- Sample Survey
- Government Statistics

Physical, Chemical, Earth and Atmospheric Sciences (10%)
- Chemistry, Chemometrics
- Geology, Geophysics
- Physics, Astronomy,Chaos
- Meteorology
- Oceanography

Agriculture (4%)
- Animal Science
- Soils and Crop Sciences
- Agricultural Economics
- Veterinary Medicine
- Food Science

Business Administration (4%)
- Finance
- Forecasting

Law (2%)

STATISTICAL CULTURE RESEARCH PROBLEMS:

DEFINITIONS OF STATISTICS AND STATISTICAL SCIENCE. Is a suitable definition of statistics (which is similar to that of McPherson (1989), p. 224) "form expectations, make observations, compare observations and expectations, continuously improve"? Is a suitable definition of statistical science "the science of analyzing data by varying conditions (probability models and estimation criteria) under which one analyzes a data set"? Note that laboratory science learns about a phenomenon by varying the experiments conducted to generate observations about the phenomenon.

EFFECTIVENESS RANKING OF STATISTICS PROGRAMS: Statistics programs in U. S. universities are usually ranked by their contributions to research in statistical methods and theory. Should they also be ranked by their effectiveness with regard to their success in adding to the U. S. work force new degree holders (bachelors, masters, doctorates) who have received education to practice statistics at the various levels we have identified? Should we regard as unsatisfactory the following current appropriate proportions being produced on the average in the U. S.

consumers (pre-calculus course)	800/10000
consumers (post-calculus course)	200/10000
appliers	100/10000
consultants	10/10000
collaborators	4/10000
theorists	2/10000

One category in which it is particularly urgent for statistics programs to increase the number of students is consumer (post-calculus) courses since this is the source which suplies candidates for all other levels of statistical practice. Desirable goals for the fraction of students in introductory courses who are taking a course with calculus prerequisite is

30%.

UNDERGRADUATE EDUCATION: Provide students with a grid of introductory courses in statistics which introduce the elegance and utility of statistical thinking, meet the needs for training at various levels of statistical literacy, are appropriate to students' scientific interests and mathematical backgrounds, and meet the goals of training all workers to become statistically literate at the consumer level, and many researchers to become statistically literate at the applier level.

The television series "Against All Odds" provides excellent supplementary material for undergraduate statistical education. An exposure to the methods and applications discussed in "Against All Odds" can be defined to be a superior grade of statistical literacy at the consumer level.

GRADUATE EDUCATION: Design graduate education in statistics to successfully provide training at each level of the practice of statistics, and which educates graduate students to have broad interests in applied, theoretical, and computational modern statistics. Students should have available courses in statistical culture which expose them to the role played by statistical methods in each of the disciplines listed in Table 1.

One of the important expected benefits of the study of statistical culture is to help the development of communication, mutual respect and cooperation between statisticians involved with various levels of practice of statistics. Graduate students in statistics come from an extreme diversity of backgrounds. The study of statistical culture would actively encourage them to communicate more with each other (as well as with their faculty) about the expertise which they should acquire as students and also during their careers. Such discussions should be part of the graduate curriculum in a first year course (which could be called Statistical Forum or Statistical Culture) which would also help students decide about whether they want a master's or doctor's degree.

STATISTICS AND RELATED FIELDS: Identify the relations between statistics and mathematics, between statistics and probability, between statistics and computing, and between statistics and the design of scientific investigations.

STATISTICAL VITALITY: How much of the current vitality of statistics derives from the availabilility of jobs in industrial statistics, biostatistics, and environmental statistics? Further, how do these areas of application compare with regard to the comparative development of the various levels of statistical practice?

THE URGENT NEED FOR MERGERS OF STATISTICIANS!

Statisticians in the United Kingdom are currently calling for a more unified less confusing public image of Statistics by merging the Royal Statistical Society and the Institute of Statisticians. Statistical Culture is the study of how statisticians of various levels can successfully merge.

If we want to successfully achieve "Viva Statistical Science" is it a prerequisite to also successfully acheive "Viva Statistical Culture"? I believe that the answer is an unequivocal yes if we take as our motto "Always remember... Statistics is Fun" (where fun can have one or more of the meanings: fun (elegant), functional (useful), functional (abstract analysis), function (graphical), function (estimation), fundamental).

References

Bodmer, W. F. (1985). "Understanding Statistics," *J. R. Statist. Soc. A*, 148,69–81.

Boroto, D.J. and Zahn, D.A. (1989). "Promoting statistics: on becoming valued and utilized" *The American Statistician*, 43, 71-72.

Cox, L. H., Voytuk, J. A., and Hart, T. A. (1989) "Trends in Mathematics and Statistics". Report of the National Research Council as part of study The Mathematical Sciences in the Year 2000.

Hahn, G.H. (1989). "Statistics aided manufacturing: a look into the future", *The American Statistician* 43, pp. 74-79.

Mann, N. R. (1988) "Why it happened in Japan and not in the U.S.", *Chance*, 1, 8-15.

McPherson, G. (1989). "The Scientists' View of Statistics–A Neglected Area," *Journal Royal Statistical Society A*, 152, 221–240.

Olkin, I. and Sacks, J. (1988) "Cross-disciplinary research in the Statistical Sciences," Institute of Mathematical Statistics.

Parzen, E. (1989) "Multi-Sample Functional Statistical Data Analysis," *Statistical Data Analysis and Inference Conference in Honor of C. R. Rao*, ed. Y. Dodge, Amsterdam: Elsevier.

Penzias, A. (1989) "Teaching Statistics to Engineers," Editorial (2 June 1989) *Science*, 244, 1025.

Scully, A. (1989) "The Relationship Between Business and Higher Education: a perspective on the 21st century", *Communications of the ACM*, 32, 1056–1061.

PROCLAMATION

City of College Station

WHEREAS, there is no future without statistics,

WHEREAS, the future of our nation requires every citizen to have statistical maturity to understand and implement decisions inevitably based on the analysis of data,

WHEREAS, students planning careers should be made aware of the importance, relevance, and beauty of statistical science,

WHEREAS, to help accomplish the above goals the week of April 23rd - 29th has been proclaimed National Science and Technology Week and Mathematics Awareness Week,

WHEREAS, to stimulate awareness of statistics as a discipline at the interface of science and mathematics, the Statistics Department of Texas A&M University is organizing a program for Statistical Science Awareness Day on April 21, 1989,

NOW THEREFORE I, Larry J. Ringer, Mayor of the City of College Station, do hereby proclaim April 21, 1989 as:

"STATISTICAL SCIENCE AWARENESS DAY"

in College Station, Texas, and urge all citizens to study the proposition that quality of life in the high tech world of the future requires each person to have some level of statistical maturity.

PASSED AND APPROVED THIS THE 13th DAY OF April, 1989.

Contributions to the Theory of Random Mappings

Bernard Harris

University of Wisconsin

Madison, Wisconsin U.S.A.

0. Preliminaries

In this report, some techniques for studying random mappings and related problems are discribed. This summary concentrates primarily on methodology developed by the author. Consequently, the work of other scientists, active in this area, will not receive extensive treatment in this report. A nonnograph is in preparation, which will give substantial treatment of the history of the subject and an extensive bibliography.

The present report will concentrate on two methods used by the author to obtain result in the theory of random mappings.

The first of these is the use of classical combinatorial enumeration methods. The second approach is the use of a "composition theorem" to construct generating functions. The later technique has wide generality, leading to many distinct results upon specialization of the parameters.

1. Introduction.

Let X_n be a finite set with $|X_n| = n$ and let T_n be the set of all mappings of X_n into X_n. If $\alpha, \beta \in T_n$, then define $(\alpha \cdot \beta)(x) = \alpha(\beta(x))$ for every $x \in X_n$. With no loss of generality, we can take $X_n = \{1, 2, \ldots, n\}$. (It will be convenient to introduce some exceptions later, for which the choice $X_n = \{0, 1, \ldots, n\}$ has some minor advantages). Clearly $|T_n| = n^n$.

Let P_{T_n} be a probability measure on the subsets of T_n. Various mathematical models are obtained by appropriate choice of P_{T_n}. When there is no risk of ambiguity, the measure will be denoted by P.

2. Representations of the mappings.

In this section we introduce two additional representations for a mapping $\alpha \in T_N$, which are useful in many applications.

First, there is a one-to-one correspondence between a class of labelled directed graphs G_n, known as functional diagraphs, and T_n, the set of mappings of X_n into X_n. This can be demonstrated as follows. Fix $\alpha \in T_n$ and let $x \in X_n$. The if $\alpha(x) = y$, draw the directed edge from x to y. Such a graph will have vertex set X_n and have exactly one edge emanating from each vertex. These graphs are in fact characterized by that property. Similarly, if a labelled graph whose vertex set in

X_n is given for which exactly one edge emanates from each vertex, define $\alpha(x)$ as the terminus of the edge leaving x for each $x \epsilon X_n$. Because of this isomorphism, we will identify each mapping with its corresponding graph and employ the same notation and terminology for both.

Another representation may be constructed as follows. Let A_n be an $n \times n$ matrix constructed as follows. If $\alpha(i) = j$, then let $a_{ij} = 1$, otherwise let $a_{ij} = 0$. Such a matrix has exactly one "one" in each row. Also, assume there is an $n \times n$ matrix of "zeros" and "ones" with exactly one "one" in each row. Then, if the "one" in row i is in column j, set $\alpha(i) = j, i = 1, 2, \ldots, n$.

The three representations, the mapping, the directed graph and the matrix can be used interchangeably.

3. Properties of Mappings.

Let $\alpha \epsilon T_n$ be a fixed mappping. For every $x \epsilon X_n$ define $x_0 = x, x_1 = \alpha(x), x_2 = \alpha(x_1) = \alpha^2(x), \ldots$. That is, in general let $x_{m+1} = \alpha(x_m) = \alpha^m(x_1) = \alpha^{m+1}(x_0)$, for all $m \geq 0$.

If for some $m \geq 0, \alpha^m(x) = y$, then y is the mth image of x; the set

$$S_\alpha(x) = \{x_0, x_1, \ldots\}$$

is the set of successors of x under α.

If for some $m \leq 0, \alpha^m(x) = y$, then y is a mth inverse of x under α. In general, $\alpha^m(x), m < 0$, may not exist or may not be unique.

Let

$$P_\alpha(x) = \overset{\circ}{\bigcup_{m=-\infty}} \{\alpha^m(x)\};$$

$P_\alpha(x)$ is called the set of predecessors of x.

If there exists an $m > 0$ such that $\alpha^m(x) = x$, then x is said to be a cyclic element under α and the set

$$C_\alpha(x) = \{x, \alpha(x), \alpha^2(x), \ldots, \alpha^{m-1}(x)\}$$

is the cycle containing x. The least such m is the length of the cycle containing x. If x is not cyclic, define $C_\alpha(x) = \phi$. The set of cyclic points under α is $C_\alpha = \bigcup_{x \epsilon X_n} C_\alpha(x)$.

If there is an $r \geq 0$ and an $s \geq 0$ such that

$$\alpha^r(x) = \alpha^s(y),$$

then x and y are equivalent under α. It is easy to see that this is an equivalence relation and the equivalence class containing x, $K_\alpha(x)$, is called the component containg x. This equivalence relationship decomposes X_n into equivalence classes, which are called the components of X_n under α. If $X_n = K_\alpha(x)$, then α is said to be connected (more precisely, the graph of α, G_α, is connected). Also it is easy to see that each component has exactly one cycle.

Fix x and consider the set $\{x, \alpha(x), \alpha^2(x), \ldots\}$. Since $x \epsilon X_n$ and $|X_n| = n$, this set can have at most n distinct elements. Hence there are $r \geq 0, s > 0$ such that $\alpha^r(x) = \alpha^{s+r}(x)$. The set $\{\alpha^r(x), \alpha^{r+1}(x), \ldots, \alpha^{r+s-1}(x)\}$ is the cycle in the component $K_\alpha(x)$.

A vertex $x \epsilon X_n$ is said to be of height m under α if m is the least non-negative integer such that $\alpha^m(x)$ is cyclic. The set of vertices of height m is called the mth-stratum of α, $S_{m,\alpha}$. Also, the height of α is defined as

$$H_\alpha = \max_m \{S_{m,\alpha} \neq \phi\}.$$

Note that $S_{0,\alpha}$ is the set of vertices cyclic under x, C_α.

The restriction of α to C_α defines a mapping, which we call the permutation induced by α. This mapping, denoted by α^*, is a permutation on a subset of X_n of cardinality $|C_\alpha|$.

Finally, we introduce the notion of the order of an element $\alpha \epsilon T_n$. Consider the set of distinct elements in $\{\alpha, \alpha^2, \ldots\}$. The cardinality of this set is the order of α. If α is a permutation, this reduces to the usual definition of the order of elements in a group. We denote this by $O(\alpha)$ and it is well-known that

$$O(\alpha) = O(\alpha^*) + \max(0, H_\alpha - 1).$$

4. Mathematical Models.

In this section, we provide illustrations of some of the commonly employed choices of P_{T_n} and the mathematical structures that they describe.

1. Let $P\{\alpha(i) = j\} = \frac{1}{n}, i = 1, 2, \ldots, n$ and let the random variables $\alpha(i)$ be mutually independent. The P_{T_n} is the measure which assigns probability n^{-n} to each mapping in T_n. We will refer to this as the symmetric case.

2. Let $P\{\alpha(i) = j\} = \frac{1}{n-1}$ if $j \neq i, P\{\alpha(i) = i\} = 0$, and let $\alpha(i)$ be mutually independent ramdom variables. Then P_{T_n} is the measure which assigns the uniform probability distribution over all mappings with no fixed points.

3. Let $P\{\alpha\} = n!^{-1}$ if α maps X_n onto X_n and 0 otherwise. Then P_{T_n} is the uniform measure over the set of permutations on X_n.

4. Let $P\{\alpha(i) = 1\} = p$ and $P\{\alpha(i) = j\} = \frac{1-p}{n-1}$, for $j \neq 1$. Also, let $\alpha(i)$ be independent random variables. Then if $p > \frac{1}{n}$, the set of mappings is known as mappings with an attracting center. If $p < \frac{1}{n}$, these are referred to as mappings with a repulsing center.

Other assignments lead to random rooted labelled trees, forests of random rooted labelled trees, random connected mappings, and so forth.

In the sequel, we restrict to the symmetric case. The other cases will be treated in the more extensive manuscript, which is in preparation.

5. Probability Distributionns for the Symmetric Model.

For this case, $P_{T_n} = n^{-n}$ for every mapping $\alpha \epsilon T_n$. We first establish theorem 1.

Theorem 1.

$$P\{|S_\alpha(x)| = k, |L_\alpha(x)| = j\} = \frac{(n-1)!}{(n-k)!} n^k \tag{5.1}$$

$1 \leq j \leq k \leq n$, where $L_\alpha(x)$ is the cycle in $K_\alpha(x)$;

$$P\{|S_\alpha(x)| = k\} = \frac{(n-1)!k}{(n-k)!n^k)} \tag{5.2}$$

$$P\{|L_\alpha(x)| = j\} = \sum_{k=j}^{n} \frac{(n-1)!}{(n-k)!n^k}, \tag{5.3}$$

also

$$E|L_\alpha(x)| = E\{|S_\alpha(x)| + 1\}/2. \tag{5.4}$$

Proof. Since the probability that $\alpha(i) = j, j = 1, 2, \ldots, n, i = 1, 2, \ldots, n$ is n^{-1} and the images of each element i are independent random variables, we have:

$$P\{|S_\alpha(x)| = k, |L_\alpha(x)| = j\} =$$

$$P(\alpha^r(x) \neq x, \alpha(x), \ldots, \alpha^{r-1}(x), 0 < r \leq k-1, \alpha^k(x) = \alpha^{k-j}(x)\}$$

$$= \frac{(n-1)!}{(n-k)!} n^k,$$

verifying (5.1); (5.3) follows trivially. To establish (5.2), one need only sum (5.1) over j, $1 \leq j \leq k$. To establish (5.4), note that $E|L_\alpha(x)| \,|\, |S_\alpha(x)| = k\} = \frac{k+1}{2}$, therefore,

$$E_{S_\alpha(x)} E\{|L_\alpha(x)| \,|\, |S_\alpha(x)) = k\} = E\{\frac{|S_\alpha(x)| + 1}{2}\}.$$

The following theorem will be repeatedly employed.

Theorem 2. The joint distribution of $|S_0(\alpha)|, |S_1(\alpha)|, \ldots, |S_{n-1}(\alpha)|$ is given by

$$P\{|S_0(\alpha)| = n_0, |S_1(\alpha)| = n_1, \ldots, |S_{n-1}(\alpha)| = n_{n-1}\}$$

$$= \frac{n!}{n_0! n_1! \ldots n_{n-1}!} n_0! n_0^{n_1} n_1^{n_2} \ldots n_{n-2}^{n_{n-1}} n^{-n}, \tag{5.5}$$

where

$$\sum_{i=0}^{n-1} n_i = n.$$

Proof. $|S_0(\alpha)| = n$ if and only if α is one-to-one and onto; hence we obtain $n! n^{-n}$, which coincides with (5.5) when $|S_0(\alpha)| = n$.

Otherwise, assume $|S_0(\alpha)| < n$. Then, $\frac{n!}{n_0! \ldots n_{n-1}!}$ is the number of ways of partitioning X_n among the various strata. The n_0 elements in $S_0(\alpha)$ can be permuted in $n_0!$ ways. Next for each stratum $S_i(\alpha)$, with $|S_i(\alpha)| = n_i$, there are $n_i^{n_i+1}$ ways for the n_{i+1} elements in $S_{i+1}(\alpha)$ to have images in $S_i(\alpha)$.

Remark. If some stratum, say $S_i(\alpha) = \phi$, then (5.5) = 0 unless $S_{i+1}(\alpha) = \cdots = S_{n-1}(\alpha) = \phi$.

Theorem 2 is basic to many items in the sequel. Marginal distributions available from Theorem 2 are the distributions of the number of cyclic points, the distribution of the height of the mapping, the distributions of the number of elements in each stratum and the order of the mapping. The following lemma will prove useful in many application of Theorem 2.

Lemma 1. For all complex z and arbitrary positive itegers q,

$$z(z+q)^{q-1} = \sum_{m=1}^{q} \sum_{\substack{l_1+\cdots+l_m=q \\ l_1,\ldots,l_m \geq 1}} \frac{q!}{l_1!\ldots,l_m!} z^{l_1} l_1^{l_2} \ldots l_{m-1}^{l_m}. \tag{5.6}$$

Proof. If $q = 1$, the conclusion holds trivially. Therefore, assume that it holds for $1, 2, \ldots, q-1, q \geq 2$. Now

$$z(z+q)^{q-1} = z \sum_{l_1=1}^{q} \binom{q-1}{l_1-1} z^{l_1-1} q^{q-l_1}$$

$$= z^q + \sum_{l_1=1}^{q-1} \binom{q}{l_1} z^{l_1} l_1 \{l_1 + (q-l_1)\}^{q-l_1-1}$$

Since $1 \leq q - l_1 \leq q-1$, the induction hypothesis applies and we get

$$z(z+q)^{q-1} = z^q + \sum_{l_1=1}^{q-1} \binom{q}{l_1} z^{l_1} \sum_{m=1}^{q-l_1} \sum_{\substack{l_2+\cdots+l_{m+1}=q-l_1 \\ l_2,\ldots,l_{m+1} \geq 1}} \frac{(q-l_1)!}{l_2!\ldots l_{m+1}!} l_1^{l_2} \ldots l_m^{l_{m+1}}$$

$$= z^q + \sum_{m=1}^{q-1} \sum_{l_1=1}^{q-m} \sum_{\substack{l_2+\cdots+l_{m+1}=q-l_1 \\ l_2,\ldots,l_{m+1} \geq 1}} \frac{q!}{l_1! l_2! \ldots l_{m+1}!} z^{l_1} l_1^{l_2} \ldots l_m^{l_{m+1}}$$

$$= z^q + \sum_{M=2}^{q} \sum_{\substack{l_1+\cdots+l_M=q \\ l_1,\ldots,l_M \geq 1}} \frac{q!}{l_1!\ldots l_M!} z^{l_1} l_1^{l_2} \ldots l_{M-1}^{l_M}.$$

Since z^q is the term for $M = 1$, the induction is complete and (5.6) is established.

We now have:

Theorem 3. The distribution of the number of cyclical elements $|C_\alpha|$ is given by

$$P\{|C_\alpha| = j\} = \frac{(n-1)!j}{(n-j)!n^j}. \tag{5.7}$$

Proof. From Theorem 2,

$$P\{|C_\alpha| = j\} = P\{|S_0(\alpha)| = j\} =$$

$$\sum \frac{n!}{j!n_1!\dots n_{n-1}!} j!\, j^{n_1} n_1^{n_2} \dots n_{n-2}^{n_{n-1}} n^{-n},$$

the sum running over all partitions of $n-j$. We rewrite this expression in terms of non-empty partitions obtaining

$$P\{|C_\alpha| = j\} = \sum \frac{n!}{j!n_1!\dots n_m!} j!\, j^{n_1} n_1^{n_2} \dots n_{m-1}^{n_m} n^{-n}, \tag{5.8}$$

the sum running over $m = 1, 2, \dots, n-1, n_1, n_2, \dots, n_m \geq 1$ with $\sum_{i=1}^{m} n_i = n-j$. A comparison of (5.8) with (5.6) show that this is related to (5.6) with q replaced by $n-j$ obtaining

$$P\{|C_\alpha| = j\} = \frac{n!}{(n-j)!} \frac{j(j+n-j)^{n-j-1}}{n^n}$$

$$= \frac{n!}{(n-j)!} \frac{j}{n^{j+1}}, j = 1, 2, \dots, n,$$

establishing Theorem 3.

Remark. Note that (5.7) and (5.2) are identical. There does not appear to be an obvious explanation for this coincidence.

Theorem 4. The probability distribution of $|P_\alpha(x)|$ is given by

$$P\{|P_\alpha(x)| = j\} = \frac{(n-1)!j^{j-2}(n-j)^{n-j}}{(n-j)!(j-1)!n^{n-1}}, j = 1, 2, \dots, n. \tag{5.9}$$

Proof. For $j > 1$, let X_{j-1} be $j-1$ specified elements of X_n; we can designate these as $x_1, x_2, \dots, x_{j-1}$. Let x be a distinguished element of X_n not in X_{j-1}. Let T_1 be the set of mappings

α in T_n with $\alpha(X_n - (X_{j-1} \cup \{x\})) = X_n - (X_{j-1} \cup \{x\})$. Define T_2 as those mappings α with $\alpha(X_{j-1}) = X_{j-1} \cup \{x\}$ and $\alpha^k x_i = x$ for some $k > 0$. Let $T^* = T_1 \cap T_2$. Then

$$P\{|P_\alpha(x)| = j\} = \begin{pmatrix} n-1 \\ j-1 \end{pmatrix} P\{\alpha \epsilon T^*\}, \text{ for } j > 1,$$

and

$$P\{\alpha \epsilon T^*\} = P\{\alpha \epsilon T_1\} P\{\alpha \epsilon T_2\}.$$

First, we have

$$P\{\alpha \epsilon T_1\} = \left(\frac{n-j}{n} \right)^{n-j}.$$

Therefore, we need to calculate $P\{\alpha \epsilon T_2\}$. This is accomplished by restricting attention to X_{j-1} and defining the mapping α' satisfying $\alpha' x_i = \alpha x_i, i = 1, 2, \ldots, j-1$ and $\alpha' x = x$. That is, α' is the restriction of α to $x_1, \ldots, x_{j-1}$ and x becomes a fixed point.

Thus

$$P\{\alpha \epsilon T_2\} = \frac{1}{n^{j-1}} \sum \frac{(j-1)!}{n_1! \ldots n_m!} 1^{n_1} n_1^{n_2} \ldots n_{m-1}^{n_m}, \tag{5.10}$$

the sum running over all non-empty partitions of $j-1$. From lemma one, the sum in (5.10) can readily be evaluated, obtaining

$$P\{\alpha \epsilon T_2\} = \frac{j^{j-2}}{n^{j-1}}$$

and hence

$$P\{|P_\alpha(x)| = j\} = \frac{j^{j-2}}{n^{j-1}} \cdot \left(\frac{n-j}{n} \right)^{n-j} \begin{pmatrix} n-1 \\ j-1 \end{pmatrix},$$

establishing (5.9), for $j > 1$.

If $j = 1$, then $\{X_n \cdot\} - \{x\}$ is mapped into $\{X_n\} - \{x\}$; there are $(n-1)^{n-1}$ such mappings, which also yields (5.9).

Trivially, we have

Remark. If $|P_\alpha(x)| = n$, then x is cyclic and α is connected. This event has probability n^{-1} by (5.9).

Corollary 1.

$$P\{|\{\alpha^{-1}(x)\}| = j\} = \binom{n}{j}\left(\frac{1}{n}\right)^j\left(1-\frac{1}{n}\right)^{n-j}. \tag{5.11}$$

6. Asymptotic Extimates in the Symmetric Case.

We now obtain the asymptotic ($n \to \infty$) probability density functions of $|S_\alpha(x)|, |L_\alpha(x)|, |C_\alpha|$. Accordingly, we establish the following theorem.

Theorem 5. The joint asymptotic ($n \to \infty$) probability density function of $\frac{|S_\alpha(\psi)|}{\sqrt{n}}, \frac{|L_\alpha(\psi)|}{\sqrt{n}}$ is given by

$$f(u,v) = e^{-u^2/2}, \quad 0 < v \le u < \infty, \tag{6.1}$$

where $u = \frac{|S_\alpha(x)|}{\sqrt{n}}, v = \frac{L_\alpha(x)}{\sqrt{n}}$.

The asymptotic ($n \to \infty$) probability density function of

$$u = |S_\alpha(x)|/\sqrt{n}$$

is

$$f(u) = ue^{-u^2/2}, \quad u > 0. \tag{6.2}$$

The asymptotic ($n \to \infty$) probability density function of

$$v = |L_\alpha(x)|/\sqrt{n}$$

is

$$f(v) = \sqrt{2\pi}(1-\Phi(v)), \quad v > 0, \tag{6.3}$$

where $\Phi(v)$ is the cumulative distribution function of the standard normal distribution. Specifically

$$\Phi(v) = \int_{-\infty}^{v}(2\pi)^{-\frac{1}{2}}e^{-x^2/2}dx.$$

Proof. In (5.1) let $k = \sqrt{n}u, l = \sqrt{n}v$ and replace the factorials using Stirling's formula. This gives

$$P\{|S_a(x)| = \sqrt{n}u, |L_a(x)| = \sqrt{n}v\} \sim \frac{n^{n-\sqrt{n}u-\frac{1}{2}}\, e^{-\sqrt{n}u}}{(n-\sqrt{n}u)^{n-\sqrt{n}u+\frac{1}{2}}}$$

$$= \frac{e^{-\sqrt{n}u}}{n(1-\frac{u}{\sqrt{n}})^{n-\sqrt{n}u+\frac{1}{2}}}$$

Expanding $\log(1-u/\sqrt{n})$ is a power series, we get

$$f(u,v) = e^{-u^2/2}, \qquad 0 < v \leq u < \infty.$$

(6.2) and (6.3) are obtained by calculating the corresponding marginal distributions.

From these asymptotic relationships, we can obtain the following corollary.

Corollary 1. The means and variances are given by

$$E\{|L_a(x)\} \sim \frac{1}{4}(2\pi n)^{\frac{1}{2}}, \qquad \sigma^2(|L_a(x)|) \sim n\left[\frac{2}{3} - \frac{2\pi}{16}\right], \tag{6.4}$$

$$E\{|S_a(x)|\} \sim \frac{1}{2}(2\pi n)^{\frac{1}{2}}, \qquad \sigma^2(|S_a(x)|) \sim n\left[2 - \frac{\pi}{2}\right], \tag{6.5}$$

and

$$E\{|C_a|\} \sim \frac{1}{2}(2\pi n)^{\frac{1}{2}}, \qquad \sigma^2(|C_a|) \sim n\left[2 - \frac{\pi}{2}\right]. \tag{6.6}$$

7. The Composition Theorem.

In this section, we give an abbreviated treatment of the composition theorem. An extensive discussion of this theorem and some generalizations of it will be treated in the future monograph.

Let S_k be the symmetric group on $\{1, 2, \ldots, k\}$. To $\gamma \epsilon S_k$, we can associate a partition $\{r_1, r_2, \ldots r_k\}$, where r_i is the number of cycles of length r. Clearly $\sum_{i=1}^{k} i r_i = k$. The k-tuple $\{r_1, r_2, \ldots, r_k\}$ will be referred to as the class of γ. A subset M_k of S_k will be called self-conjugate if and only if it is the set of all permutations in a subset of the possible classes. It is easily seen that for every $\lambda \epsilon S_k, \lambda M_k \lambda^{-1} = M_k$. Now let W_k be given self-conjugate subsets of S_k and let $w_k = |W_k|, w_0 = 1$ and let w denote the sequence $\{w_k\}_{k=0}^{\infty}$. Define

$$\phi_w(z) = \sum_{k=0}^{\infty} w_k z^k / k! \tag{7.1}$$

Let $W_{kn}, k \leq n$ be the set of all $\beta = \gamma\tau\gamma^{-1}$, where γ is a one-to-one mapping of X_k into X_n and $\tau \epsilon W_k$. Now we enumerate the set of $\alpha \epsilon T_n$ with

1. $\alpha^* \epsilon W_{kn}$ for some $k \leq n$
2. $\alpha \epsilon T_{jn}$, the set of $\alpha \epsilon T_n$ of height $\leq j$.

The number of such mappings $\alpha \epsilon T_n$ will be decided by $V_{w,j,n}$ where $V_{w,j,0} = 1$. Also, we denote $V_{w,n-1,n}$ by $V_{w,n}$.

Theorem 6. If $n \geq 0$ and $0 \leq j \leq n$,

$$V_{w,j,n} = \sum \frac{n!}{k_0!\, k_1! \ldots k_j!} w_{k_0} k_0^{k_1} k_1^{k_2} \ldots k_{j-1}^{k_j}, \tag{7.2}$$

where the sum runs over $k_0 + k_1 + \ldots + k_j = n, k_0, k_1, \ldots, k_j \geq 0$.

Proof. This is an immediate consequence of Theorem 2. The following corollary is often very useful.

Corollary 2. Let $V_{w_k,n}$ be the number of $\alpha \epsilon T_n$ with $\alpha^* \epsilon W_{kn}$, k fixed and $1 \leq k \leq n$.

Then

$$V_{w,n} = \sum_{k=1}^{n} V_{w_k,n} \tag{7.3}$$

and

$$V_{w_k,n} = \binom{n-1}{k-1} w_k n^{n-k} \tag{7.4}$$

Proof. The proof follows readily from Lemma 1.

We now define

$$V_{w,j,n}(t_0, t_1, \ldots, t_n) = \sum \frac{n!}{k_0!\, k_1! \ldots k_j!} w_{k_0} k_0^{k_1} k_1^{k_2} \ldots k_{j-1}^{k_j} t_0^{k_0} t_1^{k_1} \ldots t_j^{k_j} \tag{7.5}$$

and

$$\Psi_{w,j}(z; t_0 t_1, \ldots, t_j) = \sum_{n=0}^{\infty} V_{w,j,n}(t_0, t_1, \ldots, t_j) z^n / n!. \tag{7.6}$$

This leads to Theorem 7.

Theorem 7.

$$\Psi_{w,0}(z;t_0) = \Phi_w(zt_0) \tag{7.7}$$

and for $j \geq 1$

$$\Psi_{w,j}(z;t_0,t_1\ldots,t_j) = \Psi_{w,j-1}(z;t_0,t_1,\ldots,t_{j-2},t_{j-1}e^{zt_j}). \tag{7.8}$$

Remark. This theorem can be established in a formal algebraic sense. To obtain an equivalent analytic formula, one needs to restrict to $|zt_{j-1}e^{zt_j}| < e^{-1}$ and $\max\{|zt_0|,\ldots,|zt_j|\} < e^{-1}$. Such details are omitted here but are essential for asymptotic analysis.

Proof. Write

$$\Psi_{w,j}(z_i,t_0,t_1,\ldots t_j) = \sum_{n=0}^{\infty}\frac{z^n}{n!}\sum_{q=0}^{n}\sum_{k_0,\ldots,k_{j-1}}\frac{n!}{k_0!k_1!\ldots k_{j-1}!(n-q)!}\cdot$$

$$\cdot w_{k_0}t_0^{k_0}(k_0t_1)^{k_1}\ldots(k_{j-2}t_{j-1})^{k_{j-1}}(k_{j-1}t_j)^{n-q}$$

$$= \sum\frac{z^q}{q!}\sum_{k_1,\ldots,k_{j-1}} w_{k_0}t_0^{k_0}(k_0t_1)^{k_1}\ldots(k_{j-2}t_{j-1})^{k_{j-1}}\sum_{n=q}^{\infty}\frac{(k_{j-1}t_jz)^{n-q}}{(n-q)!}$$

$$= \sum_{q=0}^{\infty}\frac{z^q}{q!}V_{w,j-1,q}(t_0,\ldots,t_{j-2},t_{j-1}e^{zt_j}).$$

Let $\Lambda_0(z_0) = z_0$, $\Lambda_1(z_0,z_1,) = z_0e^{z_1}$ and for $j \geq 2$ let

$$\Lambda_j(z_0,z_1,\ldots,z_j) = \Lambda_{j-1}(z_0,z_1,\ldots,z_{j-2},z_{j-1}e^{z_j}).$$

Theorem 8. For $j \geq 1$, we have

$$\Lambda_j(z_0,z_1,\ldots,z_j) = z_0e^{\Lambda_{j-1}(z_1,\ldots,z_j)} \tag{7.9}$$

Proof. The conclusion is immediate for $j = 1$. Assume that the result is valid for $j-1$, $j \geq 2$. Then

$$\Lambda_j(z_0,z_1,\ldots,z_j) = \Lambda_{j-1}(z_0,z_1,\ldots,z_{j-2},z_{j-1}e^{z_j})$$

$$= z_0\exp\{\Lambda_{j-2}(z_0,z_1,\ldots,z_{j-2},z_{j-1}e^{z_j})\}$$

$$= z_0\exp\{\Lambda_{j-1}(z_1,z_2,\ldots,z_{j-1},z_j)\},$$

establishing the theorem.

We now state and prove the composition theorem.

Theorem 9.

$$\Psi_{w,j}(z; t_0, t_1, \ldots, t_j) = \Phi_w(\Lambda_j(zt_0, zt_1, \ldots, zt_j). \tag{7.10}$$

Proof. For $j = 0$ this is a consequence of Theorem 7. Also from Theorem 7,

$$\Psi_{w,j}(z; t_0, t_1, \ldots, t_j) = \Psi_{w,j-1}(z; t_0, t_1, \ldots, t_{j-2}, t_{j-1}e^{zt_j}).$$

Therefore,

$$\Psi_{w,j-1}(z; t_0, t_j, \ldots, t_{j-2}, t_{j-1}e^{zt_j}) = \Phi_w(\Lambda_{j-1}(zt_0, \ldots, zt_{j-2}, zt_{j-1}e^{zt_j})$$

$$= \Phi_w(\Lambda, (zt_0, \ldots, zt_{j-2}, zt_{j-1}, zt_j)).$$

The following corollaries can now be easily established.

Corollary 3.

$$\Psi_{w,j}(z; t_0, \ldots, t_j) = \Phi_w(\Lambda_j(zt_0, \ldots, zt_j)) \tag{7.11}$$

Corollary 4. Let $\Lambda_j(z) = \Lambda_j(z, z, \ldots, z)$. Then

$$\Psi_{w,j}(z) = \Phi_w(\Lambda_j(z)) = \sum_{n=0}^{\infty} V_{w,j,n} z^n / n!, \tag{7.12}$$

where $\Lambda_0(z) = z$ and for $j \geq 1$

$$\Lambda_j(z) = ze^{\Lambda_{j-1}(z)}. \tag{7.13}$$

Corollary 5. Let $V_{w,j,n,k}$ be the number of mappings $\alpha \epsilon T_n$ with $\alpha^* \epsilon W_{kn}$, $\alpha \epsilon T_{jn}$ and $|C_\alpha| = k$. Then

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} V_{w,j,n,k} \frac{t^{kk}_x}{n!} = \Phi_w(\Lambda_j(zt, z, \ldots, z)). \tag{7.14}$$

The composition theorem provides enumerating formulas for mappings satisfying the hypotheses of Theorem 6. For such mappings it permits enumeration by number of points, number of points on cycles, number of points in each stratum and so on. The ability to choose W_k provides the generality of the results. Illustrations follow.

Example 1 Let W_k be the set of k cycles, then the set of mappings considered is the set of connected mappings.

Example 2 If W_k be the identity mapping for $k = 1$ and $W_k = \phi, k \neq 1$, the set of mappings is the set of rooted labelled trees

REFERENCES

(1) Bernard Harris, Probability Distributions Related to Random Mappings, *Annals of Mathematical Statistics*, 31 (1960), 1045-1062.

(2) Bernard Harris and Lowell Schoenfeld, A Composition Theorem for the Enumeration of Certain Subsets of the Symmetric Semigroup, in Graph Theory and its Applications, *Academic Press*, 1970.

HANDLING UNCERTAINTY IN INPUT TO EXPECTED VALUE MODELS

Mark A. Youngren
Requirements Directorate
US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20814-2797

ABSTRACT. Due to the large number of entities and processes that must be represented, combat models at the theater level in the Army today are *expected value* models. An expected value model is deterministic -- it uses the expected value of random variables as inputs and generally uses some sort of expected value within the internal processes. The use of expected value models creates problems in the proper interpretation of their output and ways for representing the uncertainty associated with the model input and processes.

This paper suggests a method for handling uncertainty in the input data sets (which usually contain elements that are specific realizations of random processes) in situations where the outcomes of interest can be expressed in binary variables (e.g., "success" or "failure"). A theater nuclear exchange is used as an example, having many different possible outcomes determined by random processes. A method is provided for describing the space of all possible outcomes of the exchange and partitioning the space into sets of outcomes which, if used as input into a theater-level conventional simulation, are expected to lead to significantly different results. A method for sampling the most probable outcome from each set is also explained.

This approach permits the construction of an experimental plan that requires a small number of model runs, each run expected to provide a significantly different result. From these runs an estimate of the variability in the theater combat resulting from uncertainty in the input data (in this case, the impact of a nuclear exchange) can be made.

1. Introduction. Modeling large systems and processes such as combat at the theater level is difficult. The number of possible units and interactions has driven most modelers to use an *expected value* approach. An expected value model uses the expected value of random variables as inputs and generally uses some sort of expected value within the internal processes. The models are

deterministic; that is, they will yield only one set of outputs for any given set of inputs. The use of expected value models creates problems in the proper interpretation of their output and ways for representing the uncertainty associated with the model input and processes. In a recent discussion paper, Stockton [1989] provided the following example:

"A Red unit will go northwest or northeast based on whether his strength at a given point is above or below some threshold value. Let's say that the real-world probability of being above the threshold is 0.6 and, if above, he will go northwest to face a very strong Blue force armed with Supertank. If he goes northeast (probability 0.4), he faces a relatively weaker force, armed with bows and arrows. With several replications of a stochastic model, expected losses will consider both possibilities and will develop expenditures of tank ammo and arrows; with an expected value model, he will always go toward the stronger force, and no expenditures of arrows will be observed."

Stockton correctly points out that the results of an expected value model, even when provided expected value inputs, are not the expected value of the output. He suggests that the output of such a model may be a "most likely value," using his example. However, we can offer another example which illustrates that expected value models also fail to provide a "most likely" result.

Suppose in the example provided above that the Red force has a visual sensor that can see all of the Blue forces traveling together (with probability 1) if the skies are clear, and cannot see any of the Blue force if the skies are cloudy. To simplify, suppose that the skies are either clear or cloudy, and the probability that the skies are clear is 0.6. How many Blue units are detected by the Red force? The expected value is 0.6 · (100 percent of the Blue units) + 0.4 · (0 percent of the Blue units) = 60 percent of the Blue units. Expected value models will normally apply expected values, either as inputs to the model (60 percent would be an expected value for the probability of target acquisition) or internal to the processes. Note, however, that acquiring 60 percent of the Blue force is the least likely outcome, as it occurs with probability 0! Even if we chose the most likely result of 100 percent detection (which is not the way that expected value models generally handle continuous variables as opposed to choices), we run into problems.

Now let us combine the two examples. It is reasonable to suppose that if the Red force can see the Blue force, or even a large percentage of the force, it will notice that one force is armed with Supertank and the other with bows and arrows. Thus, given detection, it will engage the weaker (bows and arrows) force. If we have the model take the most likely values in the two examples, it will (1) detect 100 percent of the Blue force and (2) go northwest to engage the Blue force. Each result is by itself most likely, yet the result is the most unlikely. Even if one modeled the Red force

detection at 60 percent, the combination of a 60 percent detection (still sufficient to distinguish between Supertank and bows and arrows) and moving northwest is unlikely.

Admittedly, these examples are simplistic. Yet it is true that expected value models not only fail to yield the expected value of the output, they also fail to yield the most likely output. What, then, is the probability associated with the output of an expected value model? The answer to that question, unfortunately, is "nobody knows." This is why expected value models can yield counterintuitive, contradictory, and/or nonsensical results when initially tested. The usual approach when this occurs is to adjust input data, processes, thresholds, etc. until the model yields "reasonable" results. Hopefully this yields a model that will provide suitably realistic results with a different input data set, but there are no guarantees. We unquestionably have no way of determining the likelihood of any given output from a complex expected value model.

2. Sources of Uncertainty. There are two areas of uncertainty properly associated with an expected value model that must be handled: uncertainty in the model input, and uncertainty in the model processes.

Unfortunately, a "blessed" input data set is often regarded as certain - if we have approval for a set of numbers to be used in the study, then those numbers are the set to use to support our analysis. Excursions from the base data set for purposes of analysis will vary only a small number of data items by design; the others remain fixed. Some input data values are truly fixed: the air distance from Bremen to Munich is an example. Other values may be fixed by scenario; for example, the daylight hours vary by latitude and time of year; a scenario will fix a time and place that will in turn determine the appropriate value for daylight. Unfortunately, these scenario-driven items are often fixed arbitrarily, even when they may have an impact upon the analysis. For example, if a force is particularly vulnerable to detection by a sensor that requires daylight, you can get different results in a summer versus winter scenario (which will in turn be different than that obtained using an arbitrary number like 8 hours or 12 hours). This difference may even be apparent in studies that seemingly are not associated with detection -- ammo rates could be significantly different, for example. This is a simple, obvious example; many others, not so easily identified, exist. We must regard the input data set as a single realization of many stochastic variables. It is not always clear which realization to select for use -- averages do not always exist and may not be appropriate. Furthermore, correlations exist between sets of these data inputs; for example, selecting the most likely or expected values of cloud cover and rain independently may yield the combination of sunny with 1 inch of rain! Note that this problem exists with stochastic (Monte Carlo) models -- they also require a fixed data set that is not varied from run to run.

Uncertainty also exists in the model processes. Stochastic models generally handle this uncertainty through random number draws, although they are also subject to problems associated with correlations (separate random number draws generally require independence) and fixed values such as thresholds. The examples provided above illustrate some of the problems associated with handling process and input uncertainty within an expected value model.

3. Addressing Uncertainty in Expected Value Models. At this point, it would be nice to be able to make a statement like "the solution to this problem is easy; one simply needs to. . . ." Unfortunately, there are no simple, universal solutions to the problems associated with addressing uncertainty in expected value models. It is clear, however, that any methods that might alleviate the problem must deal with the uncertainty associated with the data input as well as the uncertainty associated with the model processes. Furthermore, the uncertainty in the input data justifies the following assertion: *executing an expected value model only once for a given data set does not provide a meaningful result.* If an expected value model is to be used to support analysis, the user *must* be prepared to execute multiple runs, varying in some meaningful fashion the input data and/or the model processes, in order to establish some measure of the uncertainty associated with the output of such a model.

Ideally, such an approach will minimize the number of runs required (because running a large expected value model can be very costly), yet provide a significantly different result from each run, thus increasing the variance across all outputs. We want to be able to describe the probability that the conditions represented in the input for each run (or conditions similar to those represented) will occur.

We have developed an approach to handling input uncertainty in theater-level expected value models in situations when the outcomes of interest can be expressed in terms of binary variables; i.e., one can describe all events as "yes" or "no," "on" or "off," etc. The particular application that will be developed deals with a theater-level tactical nuclear exchange.

Several models of conventional warfare exist at the theater level. The model used at CAA is called the Force Evaluation Model (FORCEM). Like most theater-level models and scenarios, FORCEM is a low resolution expected value model, representing combat forces at the division and higher level and time in 12-hour steps. The Nuclear Effects Model Embedded Stochastically in Simulation (NEMESIS) research at CAA (Youngren [1989]) documents an analytic model for describing the possible outcomes of a theater-level tactical nuclear exchange. The methodology described in this paper arose from the need to summarize the stochastic outcomes of the theater-level

exchange as input to FORCEM.

4. The Scenario. In a theater-level battle where nuclear weapons may be employed, the commander of the forces on a side may have an overall objective (such as stabilizing the forward line of own troops (FLOT) in the defense or achieving a breakthrough in the offense) that will necessitate the use of nuclear weapons. In order to meet this objective, the commander will specify the *defeat criteria* against each unit -- that is, the necessary degree of damage to be achieved against each unit to meet his objective. The defeat criteria will differ from unit to unit depending upon the unit mission, the posture, the equipment, etc. The criteria applied to larger units (such as divisions) will frequently focus fires on critical subordinate units. For example, the defeat criteria for a unit might be achieving a latent lethal dose (about 450 rad) against at least 50 percent of the personnel in the unit. The defeat criteria for a particular division might be to defeat at least 50 percent of the infantry units *or* at least 40 percent of the armor units in the division.

Although the effects of a tactical nuclear laydown at the theater perspective are normally described in terms of defeating divisions, tactical nuclear weapons within the theater are targeted against forces at the company and battery level. The term *subunit* (also target or target subunit) used in this paper denotes a combat organization (such as a company) that would be targeted by a nuclear weapon. The size of the subunit will depend both upon the capabilities of the weapon system used to engage the subunit and the targeting doctrine of the firer. For example, companies may be targeted close to the FLOT using small, artillery-fired weapons, while battalions may be targeted deep using missiles or air-delivered weapons. For purposes of exposition, we will refer to the low-resolution combat organizations represented in theater models such as FORCEM (usually divisions, although other forces may be represented as well) as *units*.

There are very many targetable subunits in a typical theater scenario, on the order of 10^4. As a result, there are 2^{10^4} possible outcomes that can occur in terms of the defeat or failure to defeat each subunit. Even if we look only at the defeat or failure to defeat the low resolution aggregate units represented in our theater model (usually several hundred), we still have on the order of 2^{10^2} possible outcomes. Even with sophisticated techniques and considerable confounding, classical experimental design approaches require at least one run per variable. The large amount of time and effort required to execute even a simple run of a typical theater-level expected value model prohibit more than a few model runs for any study. Classical experimental designs therefore obviously cannot be applied. Our objective is to construct a plan that minimizes the number of different input data sets (thus minimizing the number of theater-level model runs) yet fully reflects the range of possible outcomes of the theater nuclear exchange.

5. A Method for Addressing Input Uncertainty in Expected Value Models. Describing the outcome of the theater-level nuclear exchange on each unit in terms of defeat criteria allows us to define a binary variable B_i, where $B_i = 1$ if the unit is defeated; 0 otherwise. Given the assumption that the outcome is independent between units, the outcome of any exchange is simply a set of 0's and 1's with the probability that any $B_i = 1$ equal to $p_{defeat}(i)$, the probability that unit i is defeated, $i = 1, \ldots, m$. Methods for easily calculating the probability of defeat for each targetable subunit are given in Youngren [1989]. Given m units, there are 2^m possible outcomes. Clearly, if we define defeat criteria in terms of total numbers of potential nuclear targets (on the order of 10^4), there are too many outcomes to enumerate.

At the theater level, however, defeat criteria can usually be expressed in terms of divisions and a limited number of other high value targets -- on the order of at most several hundred across a theater. Each division, in turn, will have its defeat criteria established in terms of units subordinate to that division. For example, suppose that a division j has 10 battalions of infantry (engaged as battalions), 24 armored companies (engaged as companies), and 20 batteries of artillery. The defeat criteria for this division may be 50 percent of the infantry, 40 percent of the armor, or 60 percent of both, with a separate criteria for artillery (divisional and nondivisional). In terms of maneuver subunits, 5 infantry battalions or 10 armor companies must be defeated in order to defeat the division. There are $\frac{(10+24)!}{p!\,(10-p)!\,q!\,(24-q)!}$ ways of choosing p infantry battalions and q armored battalions for defeat, and *all* combinations where $p \geq 5$, $q \geq 10$, or $(p + q) \geq 60$ percent of the subunit (which can be worked out for specific values of p and q) lead to the defeat of this division. If we assume that each subunit i, $i = 1, \ldots, 34$ has a unique probability of defeat $p_{defeat}(i)$, we probably do not wish to enumerate all sets of subunits where the division is defeated and compute the joint probability (which will be the product of $p_{defeat}(i)$ for the subunits i defeated and $(1-p_{defeat}(i))$ for the subunits that are not). Fortunately, this situation is readily amenable to Monte Carlo solutions. We simply need to draw 34 binary pseudorandom numbers B_i such that each number $B_i = 1$ with probability $p_{defeat}(i)$, and let a binary variable, say D_n, equal 1 if the set of numbers B_i drawn correspond to division j being defeated, 0 otherwise. If we perform N replications of this experiment, we can estimate P[division defeated] $= \frac{1}{N}\sum_{n=1}^{N} D_n$. If we do this for each division j, then we have a probability $p_{defeat}(div\ j) \equiv$ P[division j defeated] for $j = 1, \ldots, ndiv$, where $ndiv$ = the number of divisions.

At the division level, we can define a binary variable O_j to define the outcome of the nuclear exchange with respect to division j, $j = 1, \ldots, ndiv$. $O_j = 1$ with probability $p_{defeat}(div\ j)$ if division j is defeated; 0 otherwise.

Across the theater, the theater commander will desire at least a certain percentage of units be defeated in order for the employment of nuclear weapons to be considered effective. We can define a binary function of the random variables $\underline{O}$, $\phi(\underline{O})$, such that $\phi(\underline{O}) = 1$ if the commander's objective is met; 0 otherwise. Clearly $\phi(\underline{O})$ is nondecreasing in $\underline{O}$. The function ϕ may be regarded as identical to a structure function of a coherent system in reliability theory (Barlow & Proschan [1981]); thus we can use results from coherent structure theory in our analysis of the nuclear exchange issue.

For example, if any k out of m divisions must be defeated in order for the commander's objective to be met,

$$\phi(\underline{O}) = (O_1 O_2 \cdots O_k) \amalg (O_1 O_2 \cdots O_{k-1} O_{k+1}) \amalg \cdots \amalg (O_{m-k+1} \cdots O_m),$$

for all possible subsets of size k from the m units, $1 \leq k < m$, where

$$(x_j) \amalg (x_i) \equiv 1 - (1 - x_i)(1 - x_j).$$

Furthermore, we can bound $P[\phi(\underline{O}) = 1]$ by (Barlow & Proschan [1981] p. 31):

$$\max_{1 \leq r \leq npath} \coprod_{i \in P_r} P[O_i = 1] \leq P[\phi(\underline{O}) = 1] \leq \min_{1 \leq s \leq ncut} \amalg_{i \in K_s} P[O_i = 1],$$

where P_r denotes one of the $npath = \binom{m}{k}$ possible min path sets (in this case, a min path set is any set of k units), K_s denotes one of the $ncut = \binom{m}{m-k+1}$ possible min cut sets (in this case, a min cut set is any set of m-k+1 units), and $\amalg_i X_i = 1 - \coprod_i (1 - X_i)$. If we let $p_o(i) = P[O_i = 1]$, and number the units such that $p_o(1) \leq p_o(2) \leq \cdots \leq p_o(m)$, then

$$\max_{1 \leq r \leq npath} \coprod_{i \in P_r} P[O_i = 1] = \coprod_{i=m-k+1}^{m} p_o(i); \quad \min_{1 \leq s \leq ncut} \amalg_{i \in K_s} P[O_i = 1] = \amalg_{i=1}^{m-k+1} p_o(i).$$

This example of a k out on m defeat criteria shows how we can estimate (through bounds) the probability that the commander's objective may be met. Alternatively, $P[\phi(\underline{O}) = 1]$ can be estimated using the same Monte Carlo technique used to find $P[O_j = 1]$ for each division j.

6. Partitioning the Space of All Possible Outcomes. At the theater level with a total of nt division-sized and high value targets, if we examine the nuclear exchange outcome O_j for each division (or equivalent high-value target), there are 2^{nt} possible outcomes. It may be the case that it makes a difference in the battle that follows the nuclear exchange *which* units are defeated or targets destroyed in the exchange. Or, more simply, it may be *how many* units are defeated and targets destroyed across the theater which makes a difference.

It is possible to define sets of outcomes of the nuclear exchange that, given our best judgment, we expect to have a significantly different effect on any subsequent theater-level battle (if all outcomes have approximately the same effect, then there is one set consisting of all outcomes). We choose these sets by selecting *partitions* dividing the sample space (space of all possible outcomes) into *strata* such that the following properties are met:

(1) All events within a given stratum will yield approximately the same overall theater-level outcome. As a result of this assumption, we regard all events within any given stratum as *exchangeable*.

(2) Any set of n events from n different strata are expected to yield n different theater-level outcomes. Thus, any pair of events from two different strata are *not* exchangeable.

In practice, all events within a stratum will not be truly exchangeable, and the two events to either "side" of any partition will likely lead to similar theater-level outcomes. Nevertheless, it is possible to conceive of outcome sets with different results, and we assume for all of the development below that these two properties are obeyed.

For example, suppose that there are 20 opposing divisions in a sector of combat. Our best judgment, given the tactical and operational situation, is that the defeat of at least 7 divisions out of the 20 will be required to avoid loss of territory (stabilize the FLOT--which may be the commander's objective). However, if 14 or more divisions are defeated, an opportunity occurs not merely to stabilize the FLOT but also to conduct a successful counterattack. In this case, if $O_i = 1$ if division i is defeated, $i = 1, \ldots, 20$, there are 2^{20} possible outcomes. We can partition the sample space of possible outcomes into the $\sum_{k=0}^{6} \binom{20}{k}$ outcomes where 6 or fewer divisions are defeated, the $\sum_{k=7}^{13} \binom{20}{k}$ outcomes where 7 or more but less than 14 divisions are defeated, and the $\sum_{k=14}^{20} \binom{20}{k}$ outcomes where 14 or more divisions are defeated.

The example given above involved two partitions (three strata); the number of partitions required depends on the number of significantly different theater-level outcomes that need to be represented. Selecting the partitions will require experienced judgment and possibly some experimentation with the theater model. If one is unsure about how many partitions to select, the number of strata should equal the maximum number of theater model runs you can afford.

7. Stratified Sampling from the Sample Space. Once the sample space (space of all possible outcomes) has been identified, it is possible to perform a stratified sampling from the sample space, each sample from the outcome of the nuclear exchange model forming an input vector to the theater-level conventional model. From each stratum created by our partitions, a single realization can be sampled. A random sampling approach can be used; however, since the actual likelihood of all of the events within a stratum may vary widely, we recommend using a fixed sampling scheme, in particular sampling the mode from each partition. Given the assumption of exchangeability between events within a stratum, any choice will have a roughly equivalent effect on the theater-level outcome, so any choice is valid. Using the mode allows us to compensate for the fact that the events within the stratum are only approximately exchangeable. A modal (most likely) outcome will also form a plausible input suitable for subsequent analysis. The theater-level conventional model, such as FORCEM, will be run ns times for each of the ns strata created from $ns-1$ partitions, using the outcome selected from each stratum as an input. If the second assumption that we made in selecting the partitions is met, the ns battles simulated in FORCEM using outcomes from the ns different strata should yield noticeably different results. The response surface estimated using these ns FORCEM runs should provide a better representation of the variability possible in theater-level combat where nuclear weapons are employed than a random selection of ns outcomes from the 2^{nt} outcomes possible, where nt is the number of targetable subunits in the theater.

The question naturally arises, "what if I am wrong in selecting the partitions?" Partitioning is a judgmental process; more of an art than a science. The situation in which this technique is to be used is one where many runs of the deterministic model are not possible; therefore, it is not possible to sample the results of many outputs given many different input data sets describing different nuclear exchange outcomes. As a result, we simply do our best to try and force realizations from areas of the space of all possible outcomes where we *think* that the theater-level outcome will be different. The impact of being wrong is not much different than being right. We still have another point in the theater-level outcome space that you are sampling. The fact that the nuclear exchange outcome did *not* lead to the theater-level outcome expected should be of great interest to the analysis. Either the theater model has deficiencies in correctly representing the impact of the exchange, or the theater situation is (surprisingly) robust to the exchange. If the theater outcome that you tried to create (by selecting the nuclear exchange outcome stratum) is still of interest, another run could be attempted (if time and resources permit), sampling from a more extreme point within the stratum.

8. Selecting the Most Likely Outcome (Mode) From Each Stratum. Selecting the mode from each stratum is simple and not computationally intensive. The partitions defining the stratum will establish the outcome vectors Q that fall within each stratum. Recall that $p_o(j) = P[\, O_j = 1 \,]$, and let $q_o(j) = 1 - p_o(j)$. Order the $p_o(j)$ and $q_o(j)$'s together from the largest to the smallest value. To select the mode within each partition, go from the first value ($p_o(j)$ or $q_o(j)$) and select the outcome $O_j = 1$ for each $p_o(j)$ and the outcome $O_j = 0$ for each $q_o(j)$. Continue until each target j has an outcome assigned, making sure to assign only one outcome to each target. It will be necessary to "skip" over the higher probability ($p_o(j)$ or $q_o(j)$) for some targets j in order to have a total set of outcomes fall within the partition.

This procedure can most easily be understood through an example. Suppose we have five divisional units with the following probabilities of defeat ($P[\, O_j = 1 \,]$): $p_o(1) = 0.2$, $p_o(2) = 0.25$, $p_o(3) = p_o(4) = 0.4$, $p_o(5) = 0.6$. We also have the following strata defined in terms of number of units defeated: { 0, 1 }, { 2, 3, 4 }, and { 5 }. We order our probabilities as follows: $q_o(1) = 0.8 \geq q_o(2) = 0.75 \geq p_o(5) = q_o(3) = q_o(4) = 0.6 \geq p_o(3) = p_o(4) = q_o(5) = 0.4 \geq p_o(2) = 0.25 \geq p_o(1) = 0.2$.

The first stratum must have zero or one unit defeated. Thus our mode for the first stratum is $q_o(1) \cdot q_o(2) \cdot p_o(5) \cdot q_o(3) \cdot q_o(4)$ (i.e., outcomes $O_1=0$, $O_2=0$, $O_5=1$, $O_3=0$, $O_4=0$), with a probability equal to $(0.8)(0.75)(0.6)^3 = 0.1296$. The second stratum must have two, three, or four units defeated and the mode is $q_o(1) \cdot q_o(2) \cdot p_o(5) \cdot q_o(3) \cdot p_o(4)$, with a probability equal to $(0.8)(0.75)(0.6)^2(0.4) = 0.0864$. In this case, we "skipped" outcome $O_4=0$ with probability 0.6 and selected outcome $O_4=1$ with probability 0.4 so that we would have *at least* 2 units defeated for this strata. Note that an equally likely selection would be $q_o(1) \cdot q_o(2) \cdot p_o(5) \cdot q_o(4) \cdot p_o(3)$. The third stratum must have five units defeated and the mode is $p_o(5) \cdot p_o(3) \cdot p_o(4) \cdot p_o(2) \cdot p_o(1)$, with a probability equal to $(0.6)(0.4)^2(0.25)(0.2) = 0.0048$.

9. Interpreting the Results of Conventional Runs Using Stratified Inputs. If we wish to obtain an output measure from the theater-level conventional model that we wish to average across all possible outcomes (which is the sort of thing we normally do in our simulation models), we need to construct a weighted average from the *ns* runs conducted using the theater model. The weight assigned to the output measure from each run k would be the total likelihood of all events within stratum k, $k = 1, \ldots, ns$. If it is possible to enumerate all of the possible outcomes (*nt* sufficiently small), this likelihood can be computed directly. If *nt* is too large, we can conduct a simple Monte Carlo estimation of the probability p_k that an event chosen at random falls within stratum k, $k = 1, \ldots,$

ns. This is the straightforward process of estimating the vector { $p_1, \ldots, p_{ns}$ } from a multinomial distribution.

We can return to the previous example to illustrate an exact computation of the likelihood of all events within a stratum. Recall that the strata were defined in terms of number of units defeated: { 0, 1 }, { 2, 3, 4 }, and { 5 }. The probability that 0 units are defeated is P{ 0 } = $q_o(1) \cdot q_o(2) \cdot q_o(3) \cdot q_o(4) \cdot q_o(5) = 0.0864$. There are $\binom{5}{1} = 5$ possible outcomes leading to 1 unit destroyed; they are:

$p_o(1) \cdot q_o(2) \cdot q_o(3) \cdot q_o(4) \cdot q_o(5)$, $q_o(1) \cdot p_o(2) \cdot q_o(3) \cdot q_o(4) \cdot q_o(5)$, $q_o(1) \cdot q_o(2) \cdot p_o(3) \cdot q_o(4) \cdot q_o(5)$,
$q_o(1) \cdot q_o(2) \cdot q_o(3) \cdot p_o(4) \cdot q_o(5)$, $q_o(1) \cdot q_o(2) \cdot q_o(3) \cdot q_o(4) \cdot p_o(5)$

with a total probability of $0.0216+0.0288+0.0576+0.0576+0.1296 = 0.2952$. Thus the total likelihood of the events in the first stratum is $0.0864 + 0.2952 = 0.3816$.

The calculations for P{2}, P{3}, and P{4} are messy (more combinations) but straightforward. The likelihoods are P{2} = 0.3612, P{3} = 0.2012, and P{4} = 0.0512, for a total likelihood of 0.6136. The likelihood of the third stratum is P{5} = 0.0048.

10. Adjustments. In practice, several cases may arise where it is desirable to make some adjustments to the basic model. We describe some of them here.

a. Likelihood of any realization within a strata being too small. In some cases, the total likelihood of any realization from a particular strata may be too small to justify further consideration. An example of this is the third strata ({5}) discussed in the previous paragraph. A probability of less than 0.01 is likely small enough to ignore in our theater level modeling (this threshold is, of course, a matter of judgment) In cases such as this, we may wish to simply run the conventional theater model with the modes from the more likely (in the example, the first and second) strata.

b. The modes from two strata are outcomes that are adjacent to one another. It is possible that the modes from two strata are at the boundary of their respective strata, next to the same partition, and thus adjacent to one another in terms of an ordered outcome space. An example of this is also provided in the previous paragraph, where the modes from the first two strata are adjacent to one another in terms of units defeated (one unit defeated in the first stratum and two in the second). In order to reinforce our second assumption (different results from different strata), we may wish to make a different selection from one stratum or the other in order to avoid similar results. Two possible adjustments come to mind.

(1) The first adjustment is to select the next highest likelihood from within either stratum that does not provide the same number of units defeated as does the mode. In our example, we would choose either an outcome of zero units defeated from the first stratum or three or four units defeated from the second stratum. The most likely outcome where zero units are defeated is $q_o(1) \cdot q_o(2) \cdot q_o(3) \cdot q_o(4) \cdot q_o(5) = 0.0864$. The most likely outcome where three or four units are defeated is $q_o(1) \cdot q_o(2) \cdot p_o(5) \cdot p_o(3) \cdot p_o(4) = 0.0576$. Since $0.0864 > 0.0576$, we could choose the outcome of zero units defeated from the first stratum and keep the outcome we previously computed (two units defeated) for the second stratum.

(2) The second possible adjustment is to define partitions such that there are "gaps" between the strata. In our previous example, we might define significantly different outcomes coming from zero or one units defeated, three or four defeated, and five defeated, where the outcome of two units defeated may be an ambiguous case leading to either the same result as { 0, 1 } or { 3, 4 } defeated units. This approach may be more realistic, as the "transitional cases" at the boundaries of the exhaustive strata may lead to theater outcomes that are not as clear cut as those nearer the center of any particular stratum. The only drawback to this approach is the fact that the total likelihood of drawing results from any of the strata will not equal one.

11. Repeated Exchanges. Until now, we have assumed that there is essentially only one nuclear exchange of interest. In other words, we have assumed that the nuclear weapons will be employed during a relatively small timeframe within the overall theater battle, and that the theater battle will be conventional thereafter (at least for the duration of the conflict to be simulated). However, it is possible that a scenario may call for repeated exchanges of nuclear weapons. We can handle each exchange by defining the outcomes through binary variables and stratifying the outcome space as explained above. However, constructing an experimental plan with a reasonable number of runs of the theater model becomes difficult. The difficulty rises from the total number of possible combinations of individual exchange outcomes, even if only a few strata are chosen for each exchange. For example, only three exchanges with only three significantly different outcomes (strata) predicted per exchange will lead to $3^3 = 27$ different possible outcomes after all three exchanges. It is probably too expensive to execute this many runs of a theater-level simulation model.

To handle such a situation, we begin by determining the probability of defeating each theater-level unit and partitioning the set of all possible outcomes as explained previously. We can diagram the 27 possible outcomes for our example as shown below in Figure 1. If 27 runs are too many to execute on our theater level simulation, then we must select a smaller subset of the 27 outcomes to actually use. The question is, of course, which subset do we pick? A stochastic simulation will

randomly select paths through the "tree" (Figure 1) by selecting individual exchange outcomes randomly according to their likelihoods. When a stochastic simulation is run multiple times, the paths with a high probability of occurrence will be selected multiple times and the paths with a low probability of occurrence will be selected infrequently if at all. The result is a weighted set of outcomes that can be used to estimate the distribution of the actual outcome after three exchanges. In our case, we cannot even afford to run the model once for each possible outcome, much less multiple times. However, we have the same objective of trying to determine a set of outcomes corresponding to particular paths that can be weighted to estimate the distribution of the actual outcome after three exchanges.

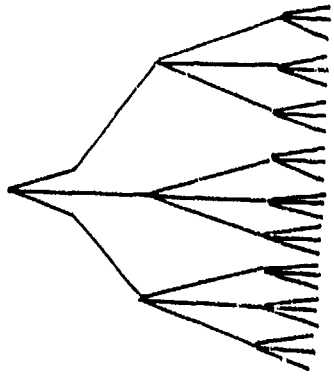


Figure 1. Possible Outcomes from Three Exchanges with Three Strata Each

Following the example diagrammed in Figure 1, let us label the strata at each exchange as high (H), medium (M), and low (L) corresponding to some exchange outcome along some measure (e.g., total units defeated). We can bound the outcome using the extreme choices at each decision point in our tree; i.e., HHH for an upper bound and LLL for a lower bound. We can also choose an intermediate outcome (MMM) in this case by choosing the intermediate result at each decision point (note that there may not always be a clearly defined "middle"). Beyond this, we need some sort of rationale for selecting particular outcomes out of the 27 possible. It is important to note that the variables are nested. For example, the middle outcome from a second strike following a high outcome from the first exchange (HM) will be different from the middle outcome from a second strike following a low outcome from the first exchange (LM), because the force strengths surviving the first exchange (and thus the subsequent theater battle before the second exchange) are significantly different.

Several approaches come to mind, both qualitative and quantitative. Qualitative approaches will choose outcomes according to the strata; for example, alternating sequences such as HML, LMH, and MLH could be chosen.

Quantitative approaches will look at the probability assigned to each stratum. For purposes of illustration, assume that the probability for the outcomes (H, M, L) are (.2, .5 .3) respectively, and that the probability for H, M, and L are identical for each of the three exchanges (in reality, this would be unlikely but it suffices for illustration). We select our runs according to their probabilities. For example, the most likely outcome will be MMM with probability $(.5)^3 = 0.125$. The next most likely are LMM, MLM, and MML with probability $(.5)^2(.3) = 0.075$, etc. We can concentrate on choosing the outcomes with the greatest likelihood (possibly in addition to the bounds HHH and LLL).

Interpreting the output becomes more difficult when we run only a subset of all possible outcome strata. In our standard experimental plan, we run all possible outcome strata and weight the result with the probability associated with the strata. If we do not make any adjustments (such as defining non-adjacent strata), the probabilities of a realization coming from a stratum will sum to 1. When we select a subset of outcome strata, the associated probabilities will not sum to 1. We recommend normalizing the probabilities associated with the outcomes selected and proceeding accordingly. An example should make this clear.

12. Repeated Exchanges -- an Example. Suppose we have three exchanges with three significantly different outcomes (strata) H, M, L with probabilities .2, .5, .3 respectively as stated previously. A possible selection scheme might be the following.

(1) Select the upper and lower bounds HHH and LLL. The associated probabilities are HHH = $(.2)^3 = 0.008$ and LLL = $(.3)^3 = 0.027$.

(2) Select the middle (qualitative) or modal (quantitative) outcome. In this case, they are the same (MMM) with probability $(.5)^3 = 0.125$.

(3) Select the next most likely outcomes LMM, MLM, and MML. The associated probabilities are equal at $(.5)^2(.3) = 0.075$. Alternatively, some type of alternating strata sequence could be used.

This forms a subset of 6 outcomes out of the 27 possible. The total probability of a realization coming from any of the 6 selected outcomes is $0.008 + 0.027 + 0.125 + (3)(0.075) = 0.385$. The normalized probabilities are therefore:

$$\text{HHH} = \frac{0.008}{0.385} = 0.021$$

$$\text{LLL} = \frac{0.027}{0.385} = 0.070$$

$$\text{MMM} = \frac{0.125}{0.385} = 0.325$$

$$\text{LMM, MLM, MML} = \frac{0.075}{0.385} = 0.195.$$

This sums to 1.001 due to rounding error.

In this example we would execute six runs of the theater-level simulation model, selecting realizations from the strata associated with each exchange as indicated above (for example, MLM would select from the middle stratum for the first and third exchange, and the lower stratum in the second). The theater-level model output associated with each realization selected can be weighted with the normalized probability of occurrence.

Note that we only account for 38.5 percent of the possible outcomes in terms of probability. As a result, our estimates made from only six runs will not be as good as those produced from a larger subset from the 27 possible.

13. Averaging the Results. To continue our example, suppose that an outcome for some particular measure from a theater conventional model such as FORCEM was 125 for a run using input from the first stratum, 75 for a run from the second stratum, and 25 for a run from the third stratum. An average value for this measure would be derived from weighting the output from a given run with the total probability of any realization coming from within the stratum. In our example, we have (125)(.3816) + (75)(.6136) + (25)(.0048) = 93.84. This value, along with the range of values produced by the three different runs (summarized perhaps with a weighted variance or other statistic), should be much more meaningful than the value obtained by running FORCEM only for some arbitrarily chosen input set for the nuclear exchange outcome.

However, a word of caution is necessary. We started with the assumption that there is more than one significantly different outcome in the theater context; in our example, there were three. A single summary measure, such as the average, does not reflect this reality. Even a sample average and variance will not inform a decisionmaker about the possible outcomes along with their associated probabilities. Since the total number of runs of the theater conventional model will be (by necessity) small, we recommend reporting *all* of the results, accompanied perhaps with a summary measure. In cases of tactical nuclear warfare, we are often concerned with relatively unlikely events (such as the exchange itself) that nevertheless have a very significant impact. Averaging obscures this fact and can lead a decisionmaker astray.

14. Summary. Using a deterministic, expected value approach to model a real-world situation such as theater-level combat poses problems in selecting input data. A deterministic simulation demands a single input data set for a model run, while the data may have to represent a process that is inherently stochastic. An example is provided in this paper. The results of a tactical nuclear exchange within a theater is inherently stochastic, driven by random events such as target acquisitions. An "average" exchange outcome cannot properly be defined; an average fails to exist in subset selection problems (for example, if 20 units out of 50 are acquired on the average, *which* 20 are to be selected as acquired in the deterministic model?) Even where averages can be defined, they fail to reflect important variations in possible outcomes that may make a difference between winning and losing the war in a theater simulation.

Ideally, a theater-level stochastic model would be used to properly reflect uncertainties inherent in the data and processes represented by the model. However, the current state of the art in hardware and software only permit us (at present) to model combat at the theater in a deterministic, low-resolution mode. Thus, we must reconcile the need to provide an input to these deterministic models with the reality of random outcomes.

If there are approximately 10^4 potential nuclear targets in a theater, there are 2^{10^4} possible outcomes that can occur in terms of the defeat or failure to defeat each potential target. Even if we look only at the defeat or failure to defeat the low resolution aggregate units represented in our theater model, we still have on the order of 2^{10^2} possible outcomes. A classical experimental design approach that requires at least one run per variable obviously cannot be applied. The challenge, then, is to construct a plan that minimizes the number of different input data sets yet fully reflects the range of possible outcomes of the theater nuclear exchange.

This paper outlines an approach to constructing such an experimental plan. We begin with the probability of defeating a potential nuclear target $p_{defeat}(i)$ and determine from that the probability of defeating the aggregate units represented in our theater model (such as divisions). We can characterize all possible outcomes of the exchange as sets of binary variables, where each binary variable reflects the defeat or failure to defeat each unit. We then partition the outcome space into strata such that outcomes from different strata lead to significantly different results in the theater battle, and all significantly different outcomes are included in some stratum. Our experimental plan consists of a nuclear exchange realization from each strata that corresponds to the most likely outcome within that stratum. The theater-level model is run using the experimental plan to determine the appropriate input data set to use to reflect the outcome of a theater nuclear exchange.

15. Directions for Future Research. The techniques outlined in this paper form only a start at trying to resolve the issue of how to handle uncertainty in input to large, complex expected value models. They are presently limited to input processes that can be summarized in a reasonable number of binary variables, where it is possible to make a judgment about the type of expected value model output given sets of similar input realizations. Nevertheless, it is a step in the right direction. At present, it is not infrequent to find studies based on a single model run per input scenario, without any estimate of the variability possible in the results obtained.

Possible future research topics include extending the techniques to processes that can be expressed in various states, the number of such states exceeding two. Better ways of estimating partitions of the sample space may also be developed. A very realistic case in many theater scenarios involves repeated realizations of random processes (in the context of the nuclear exchanges discussed in the paper, this would imply many small weapon exchanges over a relatively long period of time). At present, we have no satisfactory way of handling this situation. Robust experimental plans that can provide meaningful results over a large number of repeated realizations will be be necessary to model such scenarios.

16. References.

Stockton, G.S. (1989). *Reservations Concerning VIC as a Combat Sample Generator.* Unpublished paper.

Barlow, R.E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing - Probability Models.* To Begin With, Silver Spring, MD.

Youngren, M.A. (1989). *Nuclear Effects Model Embedded Stochastically in Simulation (NEMESIS).* Technical Paper CAA-TP-89-8, US Army Concepts Analysis Agency, Bethesda, MD.

Loading and Material Property Uncertainties in Finite Element Analyses for Orthopaedics

S.Chinchalkar* D.L.Taylor†
Sibley School of Mechanical and Aerospace Engineering
Cornell University
Ithaca , New York 14853

19 October 1989

Abstract

Stress analysis of the human femur involves uncertainties in material properties, geometry, loads and boundary conditions. It is desired to propagate these uncertainties through the Finite Element Method of stress analysis in order to obtain the distributions of stresses and displacements in the femur. This would provide better insight into bone behavior and the design of bone implants.

In particular, data from CT scans is currently used to estimate the Young's modulus of bone. The CT number at any point within the cross-section is used to estimate the apparent density at that point by means of a linear relationship. Using experimental data published by previous researchers, Young's modulus is related to apparent density.

Randomness in stresses and displacements can be studied by either a First Order-Second Moment method or by simulation. This paper compares the accuracy of FOSM with that of simulation for a simple deterministic 2-dimensional geometry. It is observed that second moment analysis can be adequate for predicting accurately the first two moments of the structural response.

Randomness in loading is much easier to analyze as compared to randomness in Young's modulus because stresses and displacements are *linear* functions of the applied loads. This paper compares the relative importance of randomness in loading to randomness in Young's modulus. Numerical experiments with random material properties show that randomness in Young's modulus has little influence on the randomness in stress when loading is also random.

*Graduate Student
†Associate Professor

1 Introduction

A "standard" Finite Element Analysis assumes all input information to be deterministic. In particular, loads, geometry, material properties and boundary conditions are assumed by the analyst, to be known precisely. Consequently, the results of such an analysis are also deterministic. In reality, there is considerable variability in this input data. This randomness affects the structural response. Frequently, designers use a 'factor of safety' to offset their lack of knowledge of the probabilistic aspects of the response.

Stochastic FEM models uncertain input information by means of random variables. The first two moments of the structural response can be obtained by a First Order Second Moment method. Such a method can provide more detailed information regarding the response as compared to the deterministic finite element method.

Finite element analysis of the femur is currently being performed assuming deterministic input, in spite of experimental evidence suggesting considerable randomness in this input data. A study of the effect of randomness in loading and material properties would help evaluate the accuracy of the deterministic solution. This paper deals with the effect of randomness in loading and material properties on a simple 2-dimensional model of the proximal femur.

2 Probabilistic Structural Analysis

Probabilistic Structural Analysis deals with analysis of structures in the presence of uncertainty. It can be used to calculate the first two moments or the distribution functions of the structural response. Structural reliability theory aims at calculating the probability of failure for structural systems. Since there are no closed-form expressions for stresses and displacements obtained by a finite element analysis, Monte Carlo simulation [Shin 72] must be used to determine the distributions of the response. Since realistic structural analysis problems tend to be computationally intensive and that detailed probabilistic information regarding the random input data is rarely available, the approximate technique of First Order Second Moment (FOSM) method is sometimes more suitable for stochastic finite element analysis.

Some of the earliest work in this field dealt with eigenvalue problems involving random media [Coll 69]. Subsequently, stochastic finite element analysis has also been applied to beams with random rigidity [Vanm 83b], turbopump blades [Nagp 87], etc.

There are several methods of modeling randomness in material properties such as Young's modulus. Vanmarcke [Vanm 83a] suggested modeling the random Young's modulus field as a spatially varying stochastic process. The Young's modulus for a finite element can then be obtained by an averaging of the stochastic field over the finite element. Liu [Liu 86] modeled the Young's modulus within an element by a linear combination of random Young's moduli at the nodes of the element. Yamazaki [Yama 88] considered the Young's moduli at centroids of finite elements as random

variables. Der Kiureghian [Kiur 88] compared the averaging method with the centroid method and observed that these two methods tend to bound the exact response variability; the centroid method usually over-estimates the variability whereas the averaging method usually under-estimates it.

3 Analysis

There is considerable variability in the input data for structural analysis problems in biomechanics. Young's modulus in bone is currently estimated using CT (Computed Tomography) scans. The grey value from these scans is used to estimate the apparent density by a linear relationship. The apparent density is related to the Young's modulus by an experimentally determined non-linear relationship. There is considerable variability in this experimental data. Therefore finite element models of the proximal femur have Young's moduli which are not deterministic. The grey values in a CT scan are used to determine the geometry. Distinction between bone and tissue is is based on a threshold which is chosen subjectively by the analyst. Hence the size of the bone being analyzed is not deterministic. Moreover the exact location and magnitudes of loads are not known precisely.

The results of a finite element analysis are affected by all these random inputs. Stochastic finite element analysis can be used to determine the amount of randomness in the response. Structural reliability can be used to determine the probability of failure. But in structural analysis of biomechanical systems, where the modeling uncertainties and approximations are high, a reliability index or a probability of failure could be very inaccurate. Modeling approximations include use of linear elastic finite element analysis instead of non-linear visco-elastic finite element analysis, isotropic material models instead of transversely isotropic material models, etc.

The present study was aimed at comparing simulation and FOSM for finite element analysis of the proximal femur. Also, the relative importance of randomness in material properties and loading was also studied. A typical *coarse* 3-D finite element model for the proximal femur contains about 300 elements and 1200 nodes. Stochastic finite element analysis of such problems is therefore too expensive. Hence it was decided to analyze a 2D plane strain model of the proximal femur instead. Deterministic analyses performed on both these models indicate that the results from a 2D model are qualitatively the same as those obtained from a 3D model.

The random Young's modulus field was modeled using the Young's modulus in each finite element as a random variable. Since the variability in Young's modulus is very high, uncorrelated fluctuations in Young's modulus in adjacent finite elements can give very unrealistic material property distributions. Therefore it was necessary to assume that the Young's moduli in different elements were correlated by a spatially varying correlation function. An exponentially decaying correlation function of the form $e^{-d/L}$ (where L is the "correlation length") was chosen because of the "intuitive" feeling that Young's moduli in elements close-by should not vary independently,

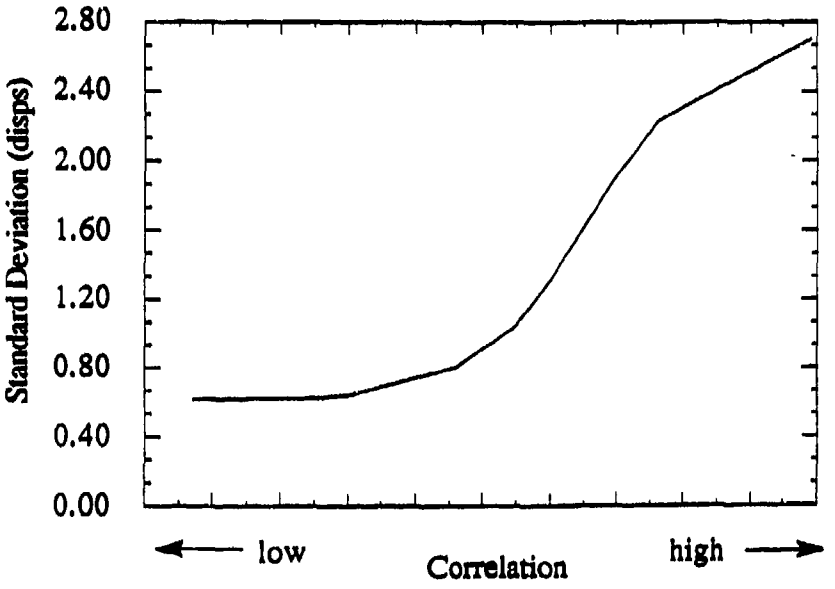


Figure 1: Variation in standard deviation of displacements

whereas Young's moduli in elements far apart could be almost uncorrelated.

Preliminary analyses showed that correlation length plays a very important role in determining the amount of randomness in the response. Figure 1 and figure 2 show the variation in the standard deviation of displacements and stresses with the correlation length for a typical plane-strain analysis. With an increase in correlation, stresses tend to become deterministic because stresses are independent of Young's moduli, provided the Young's moduli are changed uniformly by a constant factor. However the displacements in this case have maximum variability. When there is little correlation between Young's moduli, the displacements are less random but the stresses are more random. There is a considerable change in the standard deviation of the response from a fully correlated to a fully uncorrelated case. In order to obtain accurate second moments of the response, one must use a correlation function. However, the correlation function in this case must be based on experimental data.

Figure 3 shows the measured pairs of Young's modulus and apparent density [Cart 77]. The power law relationship shown is currently being used to predict the Young's modulus given apparent density. However this data cannot be used to determine a correlation function because these samples are uncorrelated and their positional data is not available. Another experimental study made by Goldstein et. al. [Gold 89] gives apparent density and Young's modulus for 8 mm specimens in the proximal and distal femur. Figure 4 compares the data presented in [Cart 77]

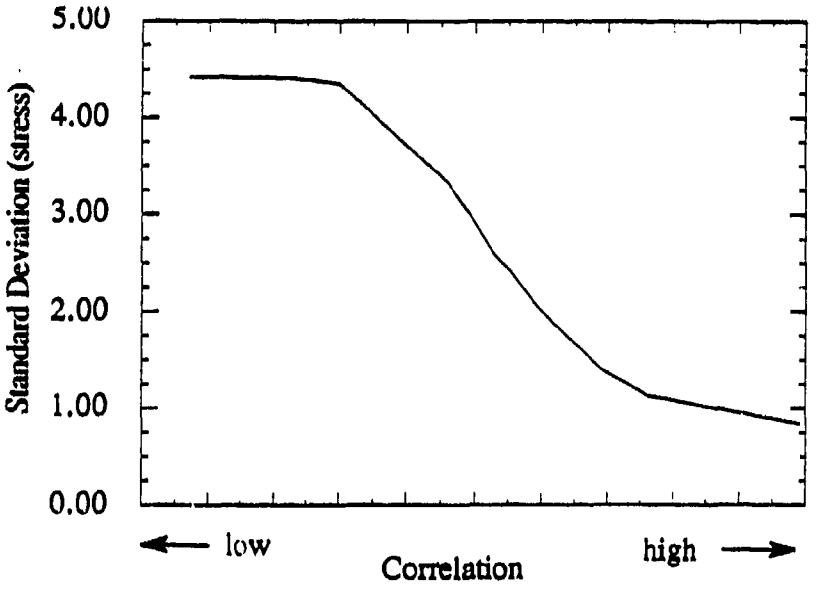


Figure 2: Variation in standard deviation of stresses

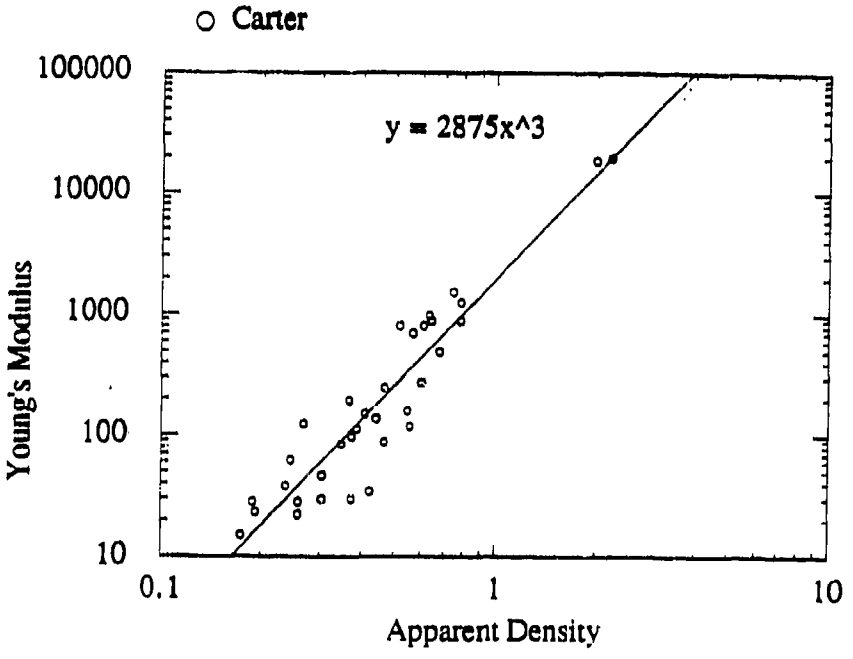


Figure 3: Young's modulus - Apparent density relationship [Cart 77]

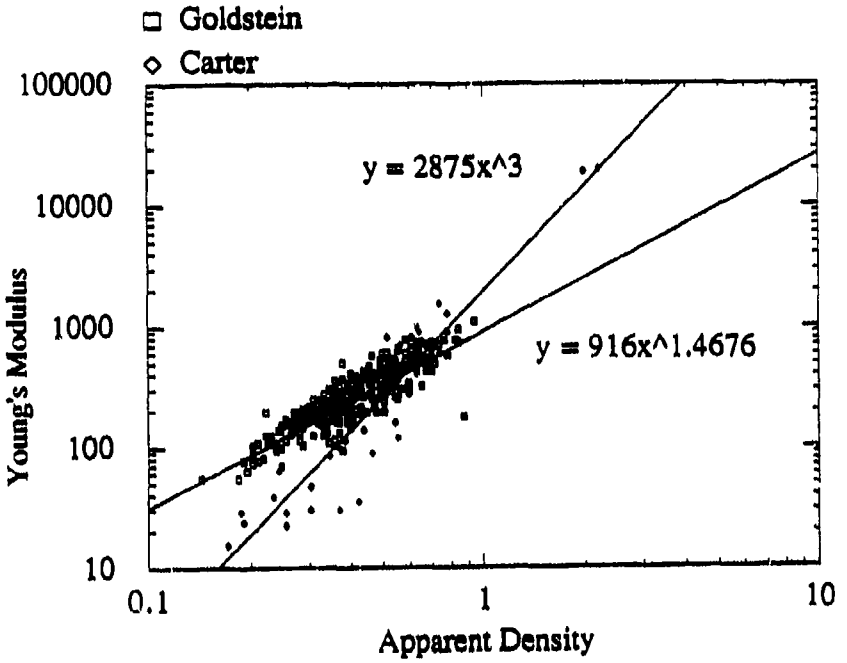


Figure 4: Comparison of data in [Cart 77] and [Gold 89]

and [Gold 89]. These two sets of data do not appear to be consistent. This can be attributed to the following :

1. The specimens in [Cart 77] came from both human as well as bovine bone.
2. [Gold 89] does not contain any data for cortical bone.
3. [Cart 77] contains both fresh and embalmed specimens from different investigators who probably performed experiments under different test conditions.

It was therefore decided to use the positional data of these specimens to estimate the correlation function, regression coefficients and variance by the method of maximum likelihood.

The following relationship was assumed to exist between the Young's modulus (E) and the apparent density (ρ)

$$\ln(E) = A + B\ln(\rho) + \epsilon \tag{1}$$

which can be written as.

$$Y = A + BX + \epsilon \tag{2}$$

where $Y = \ln(E)$ and $X = \ln(\rho)$; A and B are unknown regression coefficients and $\epsilon \sim N(0, \sigma^2)$ is a normally distributed random error. This is consistent with the linear regression on a log-log scale performed in [Cart 77]. 332 specimens were obtained from the left and right proximal and distal femurs of two cadavers ([Gold 89]).

Therefore we have

$$y_i = A + Bx_i + \epsilon_i \quad , \qquad i = 1 \text{ to } 332 \tag{3}$$

The following correlation function was chosen for the random errors ϵ_i's :

$$\mathrm{COV}[\epsilon_i, \epsilon_j] = \mathcal{E}[\epsilon_i \epsilon_j] = e^{-d/L}.\sigma^2 \tag{4}$$

where d is the Euclidean distance between the centers of specimens i and j.

The above correlation function is used with the following restrictions:

- There is no correlation between the errors ϵ_i from the proximal femur to the distal femur.
- There is no correlation between the errors ϵ_i from the left leg to the right leg.
- There is no correlation between the errors ϵ_i from one person to another.

The problem can now be stated in matrix form as follows :

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{5}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}_{n\times 2} \qquad \beta = \begin{pmatrix} A \\ B \end{pmatrix}_{2\times 1} \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n\times 1} \tag{6}$$

$$n = 332 \tag{7}$$

and

$$\mathcal{E}[\mathbf{Y}] = \mathcal{E}[\mathbf{X}\beta] + \mathcal{E}[\epsilon] = \mathbf{X}\beta \tag{8}$$

$$\mathcal{E}[(\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y} - \mathbf{X}\beta)'] = \mathcal{E}[\epsilon\epsilon'] = \sigma^2\mathbf{V} \tag{9}$$

where $\mathbf{V} = f(L)$ and L is the "correlation length". Maximum likelihood estimates for the parameters A, B, σ and L were calculated [Chin 89].

4 Results

The maximum likelihood estimates obtained are given below:

$$\begin{aligned} \hat{L} &= 16.32 \text{ mm} \\ \hat{A} &= 6.82 \\ \hat{B} &= 1.4676 \\ \hat{\sigma} &= 0.313 \end{aligned} \tag{10}$$

The relationship between E and ρ (shown in Figure 4) can now be written as

$$E = 916\rho^{1.4676}e^{\epsilon} \tag{11}$$

where $\epsilon \sim N(0, \hat{\sigma}^2)$.

Figure 5 shows the distribution of Young's modulus in the proximal femur with an implant. Titanium was chosen as the implant material and its Young's modulus (= 110 Mpa) is deterministic.

This problem was solved using both FOSM and simulation. For any function $f(\underline{x})$ (such as displacement or stress) of the random variables $\underline{x}$ (here, Young's moduli), a Taylor series expansion can be performed about the mean values of the random variables:

$$f(\underline{x}) = f(\underline{\bar{x}}) + \left(\frac{\partial f}{\partial \underline{x}}\right)^T (\underline{x} - \underline{\bar{x}}) \tag{12}$$

This yields

$$E[f(\underline{x})] = f(\underline{\bar{x}}) \tag{13}$$

$$\text{Var}[f(\underline{x})] = \left(\frac{\partial f}{\partial \underline{x}}\right)^T C_{xx} \left(\frac{\partial f}{\partial \underline{x}}\right) \tag{14}$$

where C_{xx} is the covariance matrix of the input variables and $\underline{\bar{x}}$ is the mean vector. The mean response is thus the usual deterministic response. This analysis ignores the distribution function of $\underline{x}$ and the non-linearity of $f(\underline{x})$. It is however computationally *much* faster than simulation. Simulation and FOSM results on plane-strain analyses of the proximal femur indicate that FOSM is sufficiently accurate in predicting the first two moments of the response. The error in mean and standard deviations of stresses was usually well under 5 percent. Figure 6 compares graphically the standard deviations of the stress in the inferior-superior direction obtained by these two methods. Moreover, the marginal distribution of stress at any point was very close to a Gaussian distribution. This suggests that in spite of the approximations made in FOSM analysis, FOSM can be used as a reliable alternative to simulation.

The coefficient of variation can be defined as :

$$\text{coefficient of variation} = \frac{\text{standard deviation}}{\text{mean}} \tag{15}$$

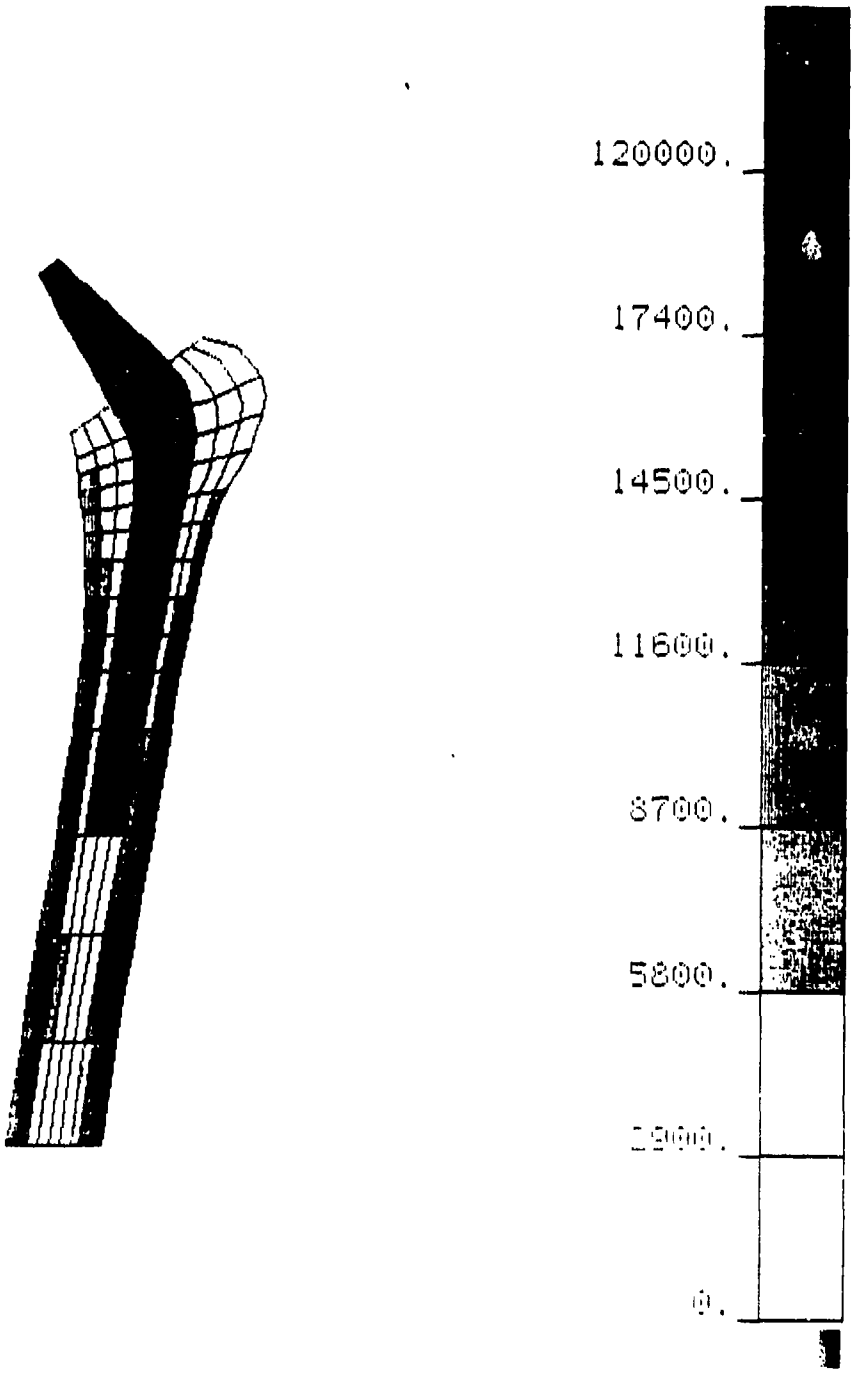


Figure 5: Distribution of the mean Young's modulus in the proximal femur

Figure 6: Comparison of standard deviation of stress obtained by simulation and FOSM

The coefficient of variation of stress as a result of randomness in Young's modulus was about a third of that of Young's modulus. This suggests that stresses are not as random as the Young's moduli.

Randomness in loading is easier to analyze because stresses and displacements are linear functions of the magnitudes of the applied loads. Thus FOSM analysis can accurately calculate the first two moments of the response. Moreover, the coefficient of variation of stresses (or displacements) is the same as the coefficient of variation of the applied loads, provided the applied loads are fully correlated. Since the applied load is not correlated to the Young's modulus, the resulting randomness in stress is dominated by the randomness in loading. Moreover if the loads are Gaussian, the resulting stresses and displacements will also be Gaussian and FOSM will again produce accurate results.

5 Conclusion

This paper studies the effect of uncertainties in material properties and loading on stresses and displacements in the proximal femur. Simulation studies showed that the approximate method of First Order Second Moment analysis can predict accurately the first two moments of the response. The resulting marginal distribution of stress was very close to being Gaussian. When the applied loads are deterministic and the Young's moduli are random, the coefficient of variation of stresses was found to be much less than that of Young's modulus. Since stresses are linear functions of the applied loads the coefficient of variation of stresses is equal to the coefficient of variation of the applied loads when the Young's moduli are deterministic. When both Young's moduli and applied loads are random, the randomness in loads dominates randomness in Young's modulus. Hence the resulting response can be predicted accurately by modeling randomness in loading alone.

6 Acknowledgements

The authors greatly appreciate the generosity of Dr. Goldstein in providing the raw data used in this paper and the guidance of Prof. Tom Santner during the initial stages of this work.

References

[Cart 77] Carter, D. R. and Hayes, W. C., "The Compressive Behavior of Bone as a Two-Phase Porous Structure", *The Journal of Bone and Joint Surgery*, V59-A, pp 954-962, 1977.

[Chin 89] Chinchalkar, S. and Taylor, D. L., "Estimation of correlation length for Young's modulus in the femoral epiphysis", technical report BE89-03, Department of Mechanical Engineering, Cornell University.

[Coll 69] Collins, J. D. and Thomson, W. T., "Eigenvalue Problem for Structural Systems with Statistical Properties", *AIAA Journal*, V7, N4, pp. 642-648, 1969.

[Kiur 88] Der Kiureghian, A. and Ke, J.-B., "The Stochastic Finite Element Method in Structural Reliability", *Probabilistic Engineering Mechanics*, V3, N2, pp. 83-91, 1988.

[Gold 89] Goldstein, S. A. et. al., "The Orthogonal Properties and Density of Human Trabecular Bone from the Major Metaphyseal Regions Utilizing Materials Testing and Computed Tomography", *Journal of Orthopaedic Research*, to be published.

[Liu 86] Liu, W. K., Belytschko, T. and Mani, A., "Random Field Finite Elements", *International Journal for Numerical Methods in Engineering*, V23, pp 1831-1845, 1986.

[Nagp 87] Nagpal, V. K. and Rubinstein, R., "Probabilistic Structural Analysis to Quantify Uncertainties Associated with Turbopump Blades", 28th *Structures, Structural Dynamics and Materials Conference*, Monterey, California, pp. 268-274, 1987.

[Shin 72] Shinozuka, M., "Monte Carlo Solution of Structural Dynamics", *International Journal of Computers and Structures*, V2, pp. 855-874, 1972.

[Vanm 83a] Vanmarcke, E., *Random Fields: Analysis and Synthesis*, MIT Press, Cambridge, Mass., 1983.

[Vanm 83b] Vanmarcke, E. and Grigoriu, M., "Stochastic Finite Element Analysis of Simple Beams", *ASCE Journal of Engineering Mechanics*, V109, N5, pp. 1203-1214, 1983.

[Yama 88] Yamazaki, F. et. al., "Neumann Expansion for Stochastic Finite Element Analysis", *ASCE Journal of Engineering Mechanics*, V114, N8, pp. 1335-1354, 1988.

NONPARAMETRIC INFERENCE UNDER MINIMAL REPAIR

Myles Hollander,* Brett Presnell,*† and Jayaram Sethuraman†

Department of Statistics
Florida State University
Tallahassee, Florida 32306-3033

FSU Technical Report No. M-818
AFOSR Technical Report No. 90-245
USARO Technical Report No. D-111

February 1990

* Research supported by Air Force Office of Scientific Research Grant AFOSR 88-0040.

† Research supported by Army Research Office Grant DAAL03-86-K-0094.

AMS (1980) subject classifcations. Primary 62N05, 90B25; secondary 62E20, 62G05, 62G10, 62G15.

Key words and phrases: Imperfect repair, life distribution, product integral.

Nonparametric Inference Under Minimal Repair

Myles Hollander
Brett Presnell
Jayaram Sethuraman
Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

Abstract

This paper summarizes the results presented at the Army Research Workshop held at Monterey, CA in October, 1989. A more detailed version will appear elsewhere.

In the age-dependent minimal repair model of Block, Borges, and Savits (1985), a system failing at age t undergoes one of two types of repair. With probability $p(t)$, a perfect repair is performed, and the system is returned to the 'good-as-new' state, while with probability $1 - p(t)$, a minimal repair is performed, and the system is repaired, but is only as good as a working system of age t. Whitaker and Samaniego (1989) propose an estimator for the system life distribution F when data are collected under this model.

Using the product integral representation of the survival function, a basic result of Block, Borges, and Savits concerning the waiting time until the first perfect repair is extended to allow for discontinuous distributions. Then using counting process techniques, the large sample theorems of Whitaker and Samaniego are extended to the whole line. These results are used to derive confidence bands for F, and to determine a sufficient condition for their applicability on the whole line. Simulation results for the bands are provided. An extension of the Wilcoxon two-sample test to the minimal repair model is also examined.

1 The Minimal Repair Model

To fix notation, let F be a life distribution, let τ_F be the upper endpoint of the support of F (possibly infinite), and let $\Lambda(t) = \int_{(0,t]}(\bar{F}(s-))^{-1}dF(s)$ be the cumulative hazard function of F, where $\bar{F} = 1 - F$.

Now, for $j = 1, \ldots, n$, let $\{X_{j,0} \equiv 0, X_{j,1}, X_{j,2}, \ldots\}$ be independent record value processes from F. These are Markov processes with $P(X_{j,k} > t \mid X_{j,0}, \ldots, X_{j,k-1}) = \bar{F}(t)/\bar{F}(X_{j,k-1})$, for $t > X_{j,k-1}$, $k \geq 1$. If $\Delta F(\tau_F) > 0$, define $X_{j,l} = \infty$ for all l larger than the first k for which $X_{j,k} = \tau_F$. In all cases we take $p(\tau_F) = 1$. These processes represent the failure ages of n systems under a "forever minimal repair" scheme.

Perfect repair is introduced into this model by the use of independent uniform random variables. This facilitates the construction of the σ-field structure (filtrations) necessary to our analysis of the model through martingale methods. Thus we let $\{U_{j,k} : 1 \leq j \leq n, k \geq 1\}$, be i.i.d. uniform r.v.'s, and define

$$\delta_{j,k} = I(U_{j,k} \leq p(X_{j,k})),$$

$$\nu_j = \inf\{k : \delta_{j,k} = 1\}.$$

Thus observing $\{(X_{j,1}, \ldots, X_{j,\nu_j}); j = 1, \ldots, n\}$, is equivalent to observing n independent copies of the age-dependent minmal repair process of Block, Borges, and Savits (BBS)(1985), each until the time of its first perfect repair.

This structure provides us with a concrete starting point for a statistical analysis of the BBS model. However, we need conditions which are sufficient to assure the finiteness of X_{j,ν_j}. Such conditions are given by the following result, which generalizes a result of BBS to the case of possibly discontinuous F. Though this generalization may not be important for modeling system failures, it will be useful to us in proving large sample results. Also, the proof of this result, which we sketch below, is more straightforward than the original proof of BBS. The reader is referred to Hollander, Proschan, and Sethuraman (1989) (HPS), for detailed proofs of this and other results in this paper.

Proposition 1 *Let* $H(t) = P(X_\nu \leq t, \nu < \infty)$. *Then*

$$\begin{aligned} \bar{H}(t) &= \Pi_{(0,t]}(1 - d\Lambda_H) \\ &= \exp\left(-\int_{(0,t]} p(s)\frac{dF^c(s)}{F(s-)}\right) \prod_{s \leq t}\left(1 - p(s)\frac{\Delta F(s)}{F(s-)}\right) \end{aligned}$$

Moreover, if either

(i) $\Delta F(\tau_F) > 0$ *(and* $p(\tau_F) = 1$*),*

or

(ii) $F(\tau_F-) = 1$ *and* $\int_0^{\tau_F} p(s)\frac{dF(s)}{F(s-)} = +\infty$,

then H *is a proper distribution function and* ν *is almost surely finite. Conversely, if* H *is a proper distribution function, then either (i) or (ii) must hold.*

Proof. (Sketch) Note that

$$\begin{aligned} \bar{H}(t) &= 1 - P(X_\nu \leq t, \nu < \infty) \\ &= 1 - \sum_{j=1}^{\infty} P(X_j \leq t, \nu = j). \end{aligned}$$

A conditioning argument shows that

$$\frac{\bar{H}(t)}{F(t)} = 1 + \sum_{j=1}^{\infty} \int \cdots \int_{0<t_1<\cdots<t_j\leq t} d\alpha(t_1)\cdots d\alpha(t_j),$$

where

$$\alpha(t) = \int_{(0,t]} (1 - p(s)) \frac{dF(s)}{F(s)}.$$

This is equivalent to

$$\frac{H(t)}{F(t)} = \prod_{(0,t]}(1 + d\alpha) = \exp(\alpha^c(t)) \prod_{s \le t}(1 + \Delta\alpha(s)),$$

where α^c is the continuous part of α and $\Delta\alpha(t)$ is the jump in α at t. Here, $\prod_{(0,t]}(1 + d\alpha)$ represents a product integral. The theory of product integration with applications in statistics is reviewed in Gill and Johansen (1987). The result follows from the last equation after some algebra. □

We will say that a pair satisfying either (i) or (ii) describes a *regular* repair scheme.

2 The Whitaker-Samaniego Estimator

In this section, we derive a martingale representation for the Whitaker-Samaniego (1989) estimator (WSE). This representation is then used in conjunction with Rebolledo's Martingale Central Limit Theorem and the techniques of Gill (1983) to derive limit theorems for the WSE.

The Basic Martingale

Define

$$N_j^*(t) = \#\{k : X_{j,k} \le t\},$$

and

$$\begin{aligned} \mathcal{F}_t &= \sigma\left(\{N_j^*(s) : s \le t, 1 \le j \le n\}\right) \\ &\quad \vee \sigma(\{U_{j,k} : k \ge 1, 1 \le j \le n\}). \end{aligned}$$

For the rest of this paper, $(\mathcal{F}_t)_{t \ge 0}$ will serve as the underlying filtration for all martingales.

Now let

$$N(t) = \#\{(j,k) : X_{j,k} \le t, k \le \nu_j, 1 \le j \le n\},$$
$$Y(t) = \#\{j : X_{j,\nu_j} \ge t, 1 \le j \le n\},$$

and

$$M(t) = N(t) - \int_{(0,t]} Y(s)\, d\Lambda(s).$$

In HPS, it is shown that M is a locally square-integrable martingale with predictable quadratic variation given by

$$\langle M \rangle(t) = \int_{(0,t]} Y(s)(1 - \Delta\Lambda(s))\, d\Lambda(s). \tag{1}$$

This provides the basic martingale structure for further analysis of the minimal repair model.

A Martingale Representation for the WSE

Assume that F is continuous and that the pair (F, p) describes a regular repair scheme. Let $X_{(k)}$ be the k^{th} ordered value of the set $\{ X_{j,k} : k \le \nu_j, 1 \le j \le n \}$, let

$$T = \min\{ X_{(k)} : Y(X_{(k)}) = 1 \},$$

and let $J(s) = I(s \le T)$. Then the Whitaker-Samaniego estimator (WSE) can be written as

$$\hat{F}(t) = \prod_{(0,t]}(1 - d\hat{\Lambda}) = \prod_{s \le t} \left(1 - \Delta\hat{\Lambda}(s)\right),$$

where

$$\hat{\Lambda}(t) = \int_{(0,t]} \frac{J(s)}{Y(s)}\, dN(s).$$

Using Duhammel's equation (Gill and Johansen, 1989), $(\hat{F} - F)/F$ can be expressed as an integral with respect to the martingale M:

$$\frac{\hat{F}(t) - F(t)}{\hat{F}(t)} = \int_0^t \frac{\hat{F}(s-)}{F(s)Y(s)}\, dM(s).$$

From this and (1) it follows $(\hat{F} - F)/F$ is itself a locally square-integrable martingale with predictable quadratic variation process given by

$$\left\langle \frac{\hat{F} - F}{F} \right\rangle = \int_{(0,t]} \left(\frac{\hat{F}(s-)}{F(s-)} \right)^2 \frac{dF(s)}{F(s)Y(s)}.$$

This quadratic variation process essentially serves to identify the covariance structure of the limiting Gaussian processes derived in the next section.

Large Sample Results

With the above representation, Rebolledo's martingale CLT and the methods of Gill(1983) yield the following result, which extends Theorem 3.3 of Whitaker and Samaniego (1989) to the whole line.

Theorem 1 *Let (F, p) describe a regular repair scheme, with F continuous. Then the following hold:*

(i) *As $n \to \infty$,*

$$\sqrt{n}\,(\hat{F} - F) \xrightarrow{\mathcal{D}} F \cdot B(C) \quad \text{in } D[0, \infty],$$

where B is Brownian motion on $[0, \infty)$, and

$$C(t) = \int_0^t \frac{dF(s)}{H(s-)\bar{F}(s)}.$$

(ii) *As* $n \to \infty$,

$$\sqrt{n}\,\frac{\hat{K}}{\hat{F}}(\hat{F} - F) \xrightarrow{\mathcal{D}} B^0(K) \quad \text{in } D[0,\infty],$$

where B^0 *is Brownian bridge on* $[0,1]$, *and* $K = C/(1+C)$.

Details of the proof of this theorem are given in HPS. We note here that the proof of (i) does not require any additional conditions beyond regularity of the repair scheme. This is in contrast with the analogous result of Gill (1983) for the Kaplan-Meier estimator in the usual censored survival data model, where some condition on the amount of censoring is needed. We will see below however, that an additional condition limiting the amount of imperfect repair is needed to assure convergence of the expression in (ii) when an estimate is substituted for $\hat{K}/\hat{F}$.

3 Applications

In this section, the asymptotic results of the last section are used to derive large sample confidence bands for F and to obtain the limiting distribution of an extension of the Mann-Whitney-Wilcoxon test statistic to the minimal repair model.

Confidence Bands

The result in part (ii) of Theorem 1 suggests confidence bands based on the distribution of the supremum of Brownian bridge. It is necessary however to estimate K/F in order to construct the bands. Let $\widehat{H}$ be the empirical cdf of the X_{j,ν_j}, and let $\widehat{K} = \hat{C}/(1+\hat{C})$, where $\hat{C}$ is defined by

$$\hat{C}(t) = \int_{(0,t]} \frac{d\hat{F}(s)}{\widehat{H}(s-)\hat{F}(s)},$$

We would like to have

$$\sqrt{n}\,\frac{\widehat{K}}{\hat{F}}(\hat{F} - F) \xrightarrow{\mathcal{D}} B^0(K) \quad \text{in } D[0,\infty], \text{ as } n \to \infty, \tag{2}$$

in order to justify asymptotic $(1-\alpha) \times 100\%$ confidence bands for F of the form

$$\hat{F} \pm \sqrt{n}\,\lambda_\alpha \hat{F}/\widehat{K},$$

where λ_α is the upper α^{th} quantile of the distribution of $\sup|B^0(t)|$.

We can show that (2) holds on $[0,\tau]$ for any $\tau < \tau_F$, but for the complete result, some additional condition seems to be needed. Using the result of Prop.1, it is shown in HPS that $\hat{K}/\hat{F}$ and $\widehat{K}/\hat{F}$ are nondecreasing and that

$$1 \le \frac{\hat{K}}{\hat{F}} \le \frac{\hat{H}_-}{\hat{F}_-} \quad \text{and} \quad 1 \le \frac{\widehat{K}}{\hat{F}} \le \frac{\widehat{H}_-}{\hat{F}_-}.$$

Using this, it can be shown that a sufficient condition for (2) is that

$$\frac{H(\tau_F-)}{F(\tau_F-)} = \int_{(0,\tau_F)} (1-p(s))\, d\Lambda(s) < \infty.$$

This condition requires that $p(t) \to 1$ as $t \uparrow \tau_F$ (at a rate sufficient for the convergence of the integral), and hence provides a limit on the amount of imperfect repair.

Simulation results for the bands computed over finite intervals (in the case of constant p) indicate that coverage probabilities are quite good for sample sizes of 50 or more. This will of course vary with the parameters of the model. Simulations were carried out with both Gamma and Weibull F, with varying shape parameters, and with various values of p, various interval lengths, and various nominal confidence levels. As an example, the following table gives the simulated coverage probabilities for nominal 95% confidence bands over the interval $[0, 4.744]$ when the underlying F is Gamma with shape parameter 2. (Note that 4.774 is the ninety-fifth percentile of Gamma(2).) More extensive tables are provided in HPS.

n	$p = .50$	$p = .25$	$p = .10$
10	.9025	.8660	.8710
20	.9270	.9125	.9187
30	.9460	.9287	.9327
50	.9515	.9398	.9395
100	.9528	.9540	.9452
200	.9515	.9517	.9495

An Extension of the Mann-Whitney-Wilcoxon Test

Using part (i) of Theorem 1, it is also possible to obtain the limiting distribution for an adaptation of the Mann-Whitney-Wilcoxon two-sample statistic to the minimal repair model. Here we assume that for $i = 1, 2$, we observe n_i BBS processes from (F_i, p_i), each until its first perfect repair. In general we wish to test the null hypothesis $H_0: F_1 = F_2$, with typical one-sided alternatives specifying $\int F_1\, dF_2 > 1/2$, and two-sided alternatives specifying $\int F_1\, dF_2 \neq 1/2$.

A statistic analogous to the Mann-Whitney form of the Wilcoxon two-sample test statistic is W, as given by

$$\begin{aligned} W &= \int \hat{F}_1\, d\hat{F}_2 \\ &= \sum_{\Delta N_2(s)>0} \hat{F}_1(s)\hat{F}_2(s-)\frac{\Delta N_2(s)}{Y_2(s)}, \end{aligned}$$

where $\hat{F}_i$ is the WSE, $\Delta N_i(s)$ is the number of failures at age s, and $Y(s)$ is the number of items at risk at age s in the i^{th} sample. This statistic is a natural estimator of $\int F_1\, dF_2 = P(X_1 \leq X_2)$, where X_1 and X_2 are independent random variables, with $X_i \sim F_i$. Assuming continuous distributions, $P(X_1 \leq X_2) = 1/2$ under H_0, and in the one-sided case, significantly large values of W provide evidence against H_0 in the direction

of $\int F_1\,dF_2 > 1/2$. For large sample sizes, we have the following result, which is proven in HPS:

Theorem 2 *If F_1 and F_2 are continuous, and the pairs (F_1, p_1) and (F_2, p_2) describe regular repair schemes, and if $n_1, n_2 \to \infty$ in such a way that $\frac{n_1}{n_1+n_2} \to \lambda$, $0 < \lambda < 1$, then*

$$\sqrt{n_1+n_2}\left[W - \int F_1\,dF_2\right] \xrightarrow{\mathcal{D}} N\left(0\,,\ \frac{1}{\lambda}\sigma_1^2 + \frac{1}{1-\lambda}\sigma_2^2\right), \tag{3}$$

where

$$\sigma_1^2 = 2\int_0^\infty \int_t^\infty F_1(s)F_1(t)C_1(t)\,dF_2(s)dF_2(t),$$
$$\sigma_2^2 = 2\int_0^\infty \int_t^\infty F_2(s)F_2(t)C_2(t)\,dF_1(s)dF_1(t).$$

Under the null hypothesis, $H_0 : F_1 = F = F_2$,

$$\sigma_1^2 = 2\int_0^\infty F(t)C_1(t)\left(\int_t^\infty F(s)\,dF(s)\right)dF(t) = \frac{1}{4}\int_0^\infty \frac{F^3(s)}{H_1(s-)}\,dF(s).$$

For purposes of testing the null hypothesis in the large sample case, we thus propose referring the test statistic

$$Z = \left(W - \frac{1}{2}\right) \Big/ \left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}\right)^{\frac{1}{2}}$$

to a standard normal distribution, where

$$\hat{\sigma}_i^2 = \frac{1}{4}\int_0^\infty \frac{\hat{F}_i^3(s)}{\hat{H}_i(s-)}\,d\hat{F}_i(s) = \frac{1}{4}\sum_{\Delta N_i(s)>0} \frac{n\hat{F}_i^3(s)\hat{F}_i(s-)\Delta N_i(s)}{Y_i^2(s)},$$

and H_i is the empirical distribution of the perfect repair ages in the i^{th} sample.

It is shown in HPS that the σ_i are consistent, which justifies the use of this test. If the p_i are constants (see Brown-Proschan (1983)), the above expressions simplify greatly under H_0. If $F_1 = F_2 = F$, then $H_i = F^{p_i}$, and the asymptotic variance in (3) reduces to

$$\frac{1}{\lambda}\sigma_1^2 + \frac{1}{1-\lambda}\sigma_2^2 = \frac{1}{\lambda}\left(\frac{1}{4(4-p_1)}\right) + \frac{1}{1-\lambda}\left(\frac{1}{4(4-p_2)}\right).$$

The p_i's are of course consistently estimated by their MLE's, $\hat{p}_i$, the ratio of n_i to the total number of failures in the i^{th} sample, and for large samples, the statistic Z', given by

$$Z' = \left(W - \frac{1}{2}\right) \Big/ \left[\frac{1}{4n_1(4-\hat{p}_1)} + \frac{1}{4n_2(4-\hat{p}_2)}\right]^{1/2},$$

can be referred to a standard normal distribution in order to test the null hypothesis. Note also that if $p_1 = p_2 = 1$, then we are in the usual i.i.d. two-sample model, the WSE's reduce to the empirical c.d.f.'s, and W is just a multiple of the Mann-Whitney form of the Wilcoxon rank-sum statistic. In this case, the above results yield

$$\left(W - \frac{1}{2}\right) \Big/ \left[\frac{1}{12}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{1/2} \xrightarrow{\mathcal{D}} N(0,1),$$

in agreement with the usual results for the Mann-Whitney-Wilcoxon test.

REFERENCES

BLOCK, H. W., BORGES, W. S., and SAVITS, T. H. (1985). Age-dependent minimal repair. *J. Appl. Prob.* **22** 370–385.

BROWN, M. and PROSCHAN, F. (1983). Imperfect repair. *J. Appl. Prob.* **20** 851–859.

GILL, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. and JOHANSEN, S. (1987). Product integrals and counting processes. CWI Report MS-R8707, Centrum voor Wiskunde en Informatica, Amsterdam. To appear in *Ann. Statist.*

HOLLANDER, M., PRESNELL, B., and SETHURAMAN, J. (1989). Nonparametric Methods for Imperfect Repair Models. FSU Technical Report No. M-817, Department of Statistics, Florida State University, Tallahassee, Florida.

WHITAKER, L. R. and SAMANIEGO, F. J. (1989). Estimating the reliability of systems subject to imperfect repair. *J. Amer. Statist. Assoc.* **84** 301–309.

THE APPLICATION OF A COMPOSITE DESIGN TO TEST A COMBAT SIMULATION MODEL

Carl B. Bates
US Army Concepts Analysis Agency
Bethesda, Maryland 20814-2797

ABSTRACT. A study is to be performed that involves the determination of a mix of target acquisition systems that yields an improved capability at a lesser cost. A primary candidate for the combat simulation is a two-sided deterministic division-level ground combat model. Before the model could be used in the study, the model had to be tested to determine its capability to evaluate the combat effectiveness of mixes of target acquisition systems. The test involved four factors, one qualitative and three quantitative factors. Time constraints limited the number of simulations to 30 runs. A composite design is presented, its application is illustrated, and its efficiency is discussed.

1. INTRODUCTION. The test was to assess the sensitivity of model output to specified changes in input values for the four selected input factors. The four factors are:

TYP - Type of sensor,

FRC - The fraction of target elements for which the sensor has both coverage and line-of-sight,

TIM - The time, in minutes, that a sensor spends processing and reporting a target it has detected,

NUM - The total number of sensors employed in a model run.

Two types (A and B) of sensors were to be evaluated. Three values were ultimately selected for each of the three quantitative factors. The minimum and maximum from operational performance were taken as the lower and upper values. A "middle" value was then added. The values are:

FRC - 0.1, 0.5, 0.9,

TIM - 0, 5, 10.

NUM - 5, 15, 25.

This gave a 2x3x3x3 full design. Time constraints, however, would permit only 30 runs for the complete test.

2. EXPERIMENTAL DESIGN. Therefore, the objective is to develop an experimental design with not more than 30 design points. The design should permit assessment of a full second-order model in the three quantitative factors. Draper and John (1988) discuss response surface designs for quantitative and qualitative variables. They give some first and second-order designs for 2^k factorials and 2^{k-p} fractional factorials. The decision was made, however, that a single model involving TYP had no advantage over two models, one for each of the two types of sensors. Now the problem is to develop a response surface design (one of each sensor type) for the three three-level quantitative factors.

Let the three variables X_1, X_2, and X_3 represent the three quantitative factors. The second-order model we wish to investigate is:

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_{11}X_1^2+\beta_{22}X_2^2+\beta_{33}X_3^2+\beta_{12}X_1X_2+\beta_{13}X_1X_3+\beta_{23}X_2X_3+e$$

The 27 design points of the full 3x3x3 design are shown in Figure 1. The low, middle, and high values of the three variables are denoted by "0", "1", and "2", respectively. The eight corner points, (000), (200), (020), (220), (002), (202), (022), and (222), would be a full 2^3 design if there were no middle values. If these eight points are augmented with the center point (111) and the six center points of each plane, (211), (011), (101), (121), (110), and (112), we have a design similar to a central composite design. The design is given in Table 1 and illustrated in Figure 2. Box and Wilson (1951) introduced the concepts of composite designs. Myers (1971) and Box and Draper (1987) discuss second-order composite designs. Myers, Khuri, and Carter (1989) discuss recent and current response surface methodology research.

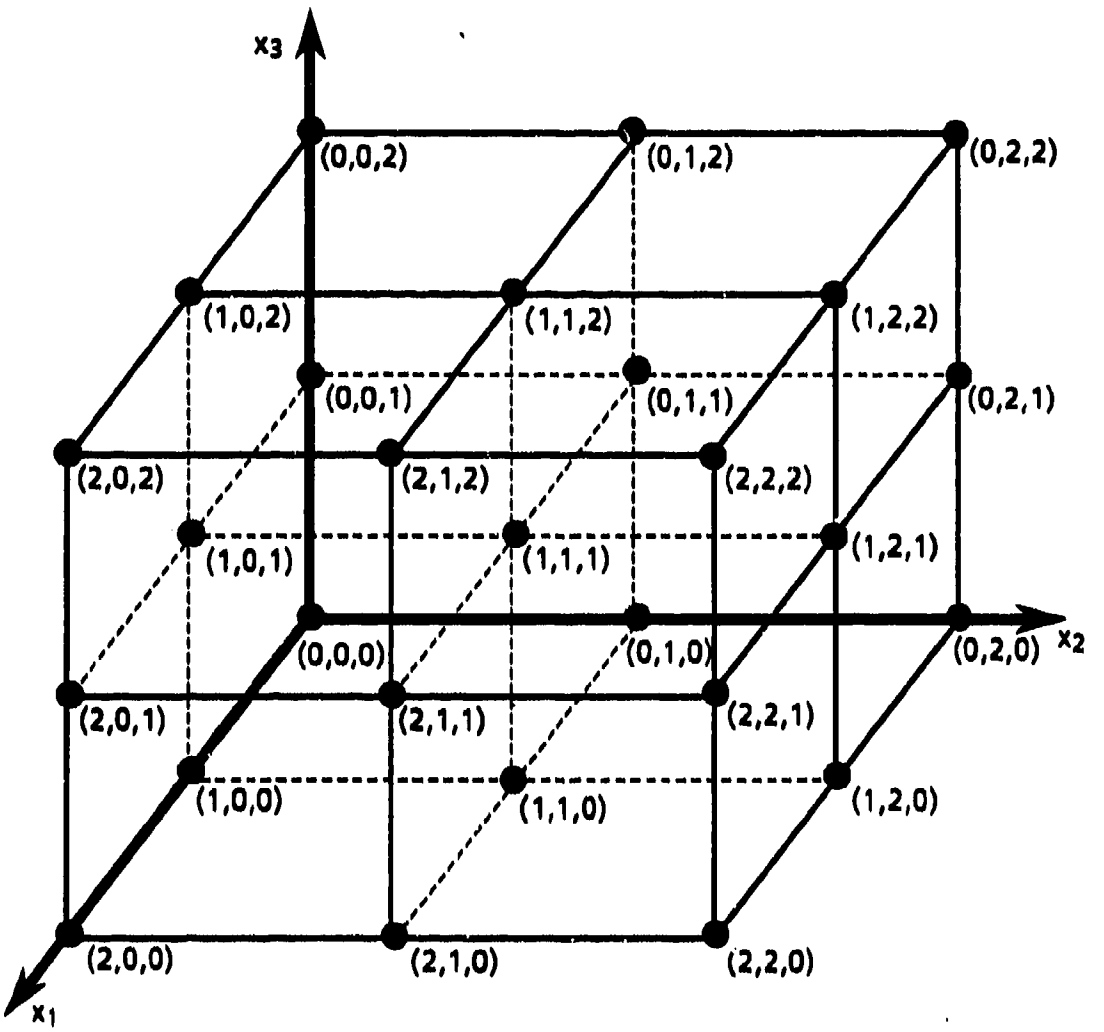


Figure 1. 3X3X3 Design

Table 1. Three-variable Composite Design

Run #	X_1	X_2	X_3	
1	0	0	0	
2	2	0	0	
3	0	2	0	
4	2	2	0	Corners
5	0	0	2	
6	2	0	2	
7	0	2	2	
8	2	2	2	
9	2	1	1	
10	0	1	1	
11	1	0	1	Star
12	1	2	1	
13	1	1	0	
14	1	1	2	
15	1	1	1	Center

A three-variable central composite design is given in Table 2. The literature on central composite designs discusses determining the value of α to yield orthogonal designs. The value of α is the length of the axial points shown in Figure 3.

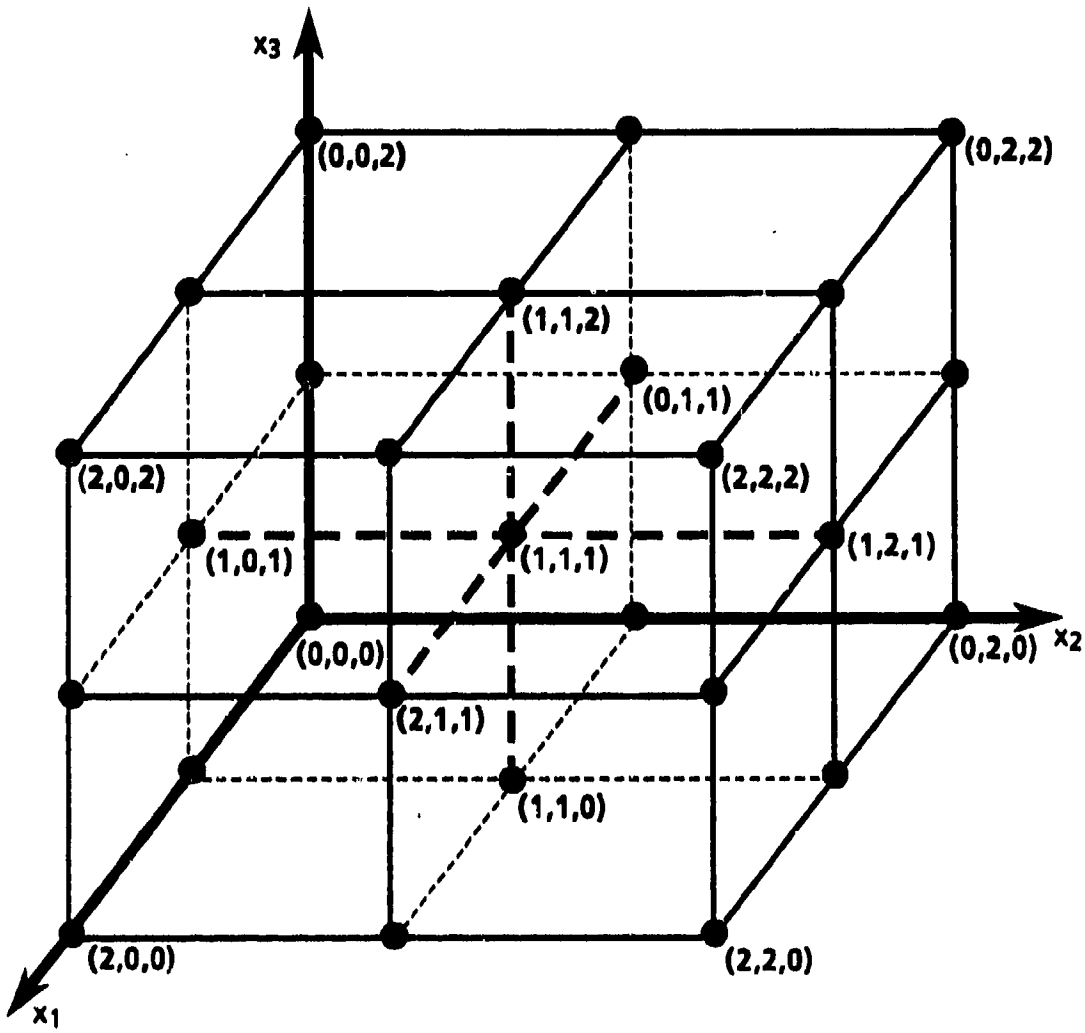


Figure 2. Composite Design

Table 2. Three-variable Central Composite Design

Run #	X_1	X_2	X_3	
1	-1	-1	-1	
2	-1	-1	1	
3	-1	1	-1	
4	-1	1	1	2^3 factoral
5	1	-1	-1	
6	1	-1	1	
7	1	1	-1	
8	1	1	1	
9	-a	0	0	
10	+a	0	0	
11	0	-a	0	Axial
12	0	+a	0	
13	0	0	-a	
14	0	0	+a	
15	0	0	0	Center

3. DESIGN EFFICIENCY. Myers (1971) discusses the efficiency of central composite designs (ccd) and shows that a three variable orthogonal ccd is as efficient as a 3^3 factorial design for estimating the mixed quadratic coefficients. The results, however, apply to only orthogonal ccd and do not apply to the restrained composite design in Table 1.

Because no information could be found on the efficiency of the restrained composite design, a cursory evaluation was made of the design. ACED, Algorithms for the Construction of Experimental Designs, developed by Welch (1985) was used for the evaluation. Welch (1984) generalizes Mitchell's DETMAX algorithm and discusses ACED. ACED has four optimality criteria, D Optimality (DO), Average Variance of the Response Estimates (AV),

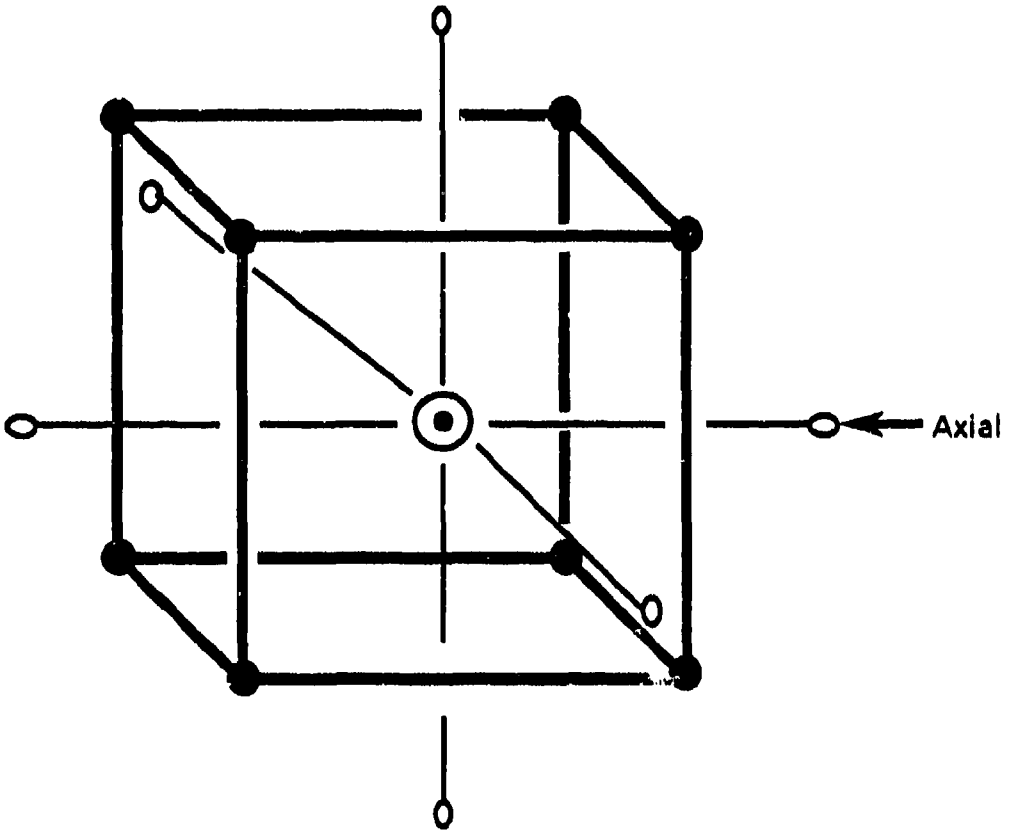


Figure 3. Central Composite Design

Maximum Variance of the Response Estimators (MV), and Average Mean Squared Error of the Response Estimators (AM). AM was selected as the evaluation criterion because it provided a robust balance between variance and bias. The AM criterion is discussed in Welch (1983).

The variances of the parameters estimates (bs) of the second-order model are:

$V(b_0) = 12.0$

$V(b_i) = 28.6$

$V(b_{ii}) = 5.8$

$V(b_{ij}) = 1.9$

The variance efficiency is 99.6% and the bias efficiency is 91.6%.

Since these efficiencies were considered acceptable and time constraints precluded further evaluation or design development, the composite design in Table 1 was employed.

4. APPLICATION. The model was exercised for each sensor type for each of the 15 design points in Table 1. Several output variables were extracted and analyzed. Testing was performed at the 0.05-level of significance. One data set, Red personnel losses, is shown in Table 3. The significant model was considered to be:

$$Y = 941.2 + 1771.1X_1 + 483.5X_3 - 9.5x_3^2 - 196.3X_1X_3$$

The unadjusted R^2 was 0.90. The residuals ($y_i - Y_i$) ranged from -743 to 568. The observed and the predicted values are shown in Figure 4. The confidence intervals on Y ranged from $\pm$481 to $\pm$701.

Table 3. Red Personnel Losses with Sensor A

Run #	X_1	X_2	X_3	y
1	0.1	0	5	1471
2	0.9	0	5	2333
3	0.1	10	5	919
4	0.9	10	5	1615
5	0.1	0	25	4313
6	0.9	0	25	2596
7	0.1	10	25	5159
8	0.9	10	25	2153
9	0.9	5	15	2670
10	0.1	5	15	4201
11	0.5	0	15	4038
12	0.5	10	15	2835
13	0.5	5	5	1823
14	0.5	5	25	3858
15	0.5	5	15	4146

The analysis results of this output variable is shown only to illustrate application of the composite design, not to illustrate goodness of the final fit.

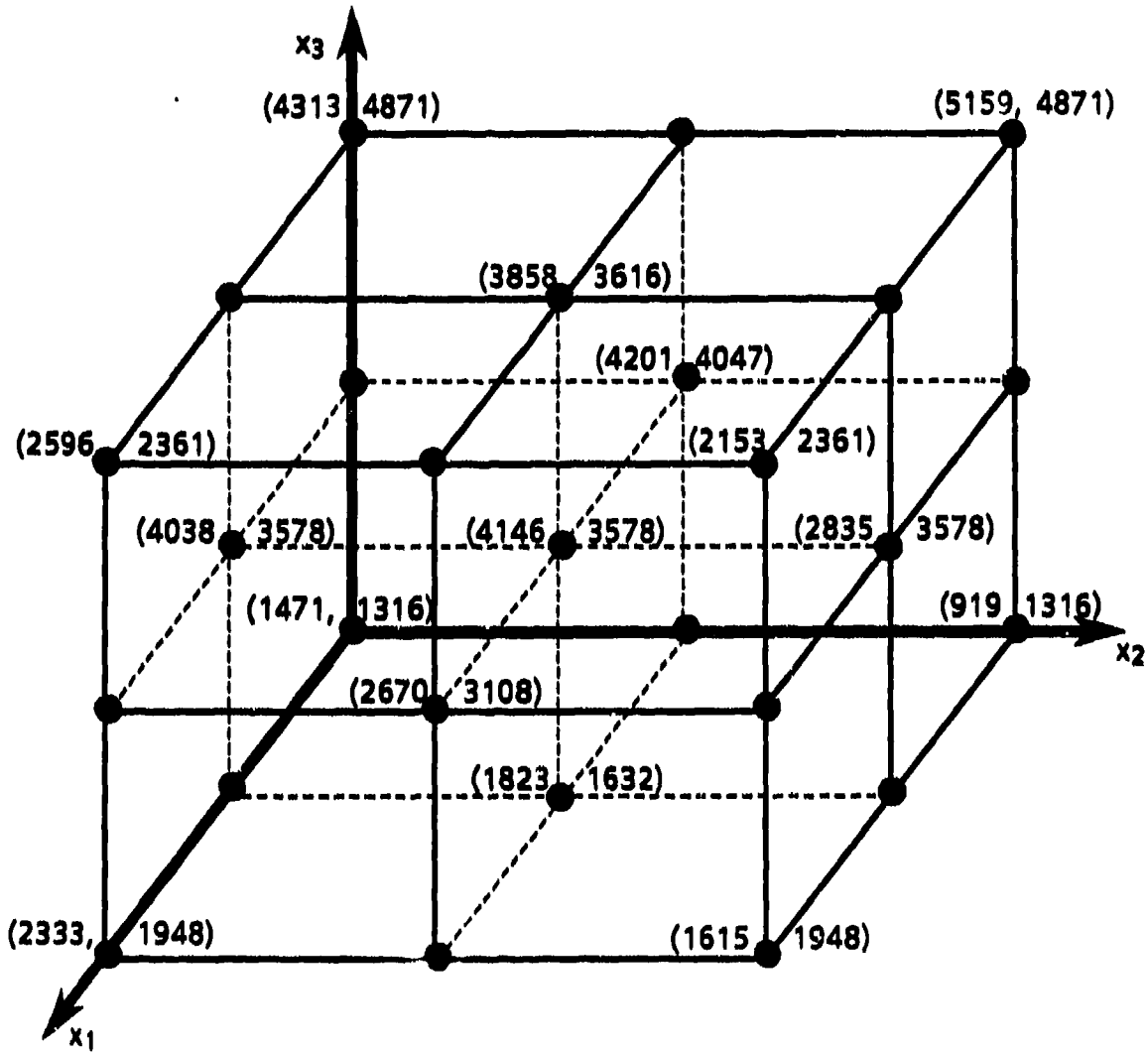


Figure 4. Observed and Predicted (y,Y) Values

5. SUMMARY. The 12-point Box-Behnken design which is the complement of the 15-point composite design used was not considered. It may have provided a more efficient design. Also not considered was shortening the six axial points to give five levels for each of the variables. This may, too, have been a superior design to the design employed. The 15-point composite design employed was considered to be appropriate for the purpose of evaluating a second-order model.

REFERENCES

1. Box, G. E. P. and Behnken, D. W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables," Technometrics, Vol. 2, pp. 455-475.

2. Box, G. E. P. and Draper, N. R. (1987), "Empirical Model-Building and Response Surfaces," John Wily & Sons, Inc., New York.

3. Box, G. E. P. and Wilson, K. B. (1951), "On the Experimental Attainment of Optimum Conditions," Journal of Royal Statistics, Series B, Vol. 13, pp. 1-45.

4. Draper, N. R. and John, J. A. (1988), "Response-Surface Designs for Quantitative and Qualitative Variables," Technometrics, Vol. 30, pp. 423-428.

5. Myers, R. H. (1971), "Response Surface Methodology," Allyn and Bacon, Inc., Boston.

6. Myers, R. H., Khuri, A. I., and Carter, W. H. (1989), "Response Surface Methodology: 1966-1988," Technometrics, Vol. 31, pp. 137-157.

7. Welch, W. J. (1983), "A Mean Squared Error Criterion for the Design of Experiments," Biometrika, Vol. 70, pp. 205-213.

8. Welch, W. J. (1984), "Computer-Aided Design of Experiments for Response Estimation," Technometrics, Vol. 26, pp. 217-224.

9. Welch, W. J. (1985), "ACED Algorithms for the Construction of Experimental Designs, User's Guide."

Distribution Theory for Variance Components Estimation Diagnostics

Dr. Jock O. Grynovicki
U.S. Army Laboratory Command
Human Engineering Laboratory
Aberdeen Proving Ground, Maryland

Dr. John W. Green
Department of Mathematics
University of Delaware
Newark, Delaware

Abstract

Distribution theory is developed for diagnostics used to investigate variance component estimates and model assumptions in mixed or random models. Estimation of variance components in a given model is the equivalent of estimation of certain linear functions thereof. Each such linear function is realized as an average of natural sample covariances, that may be independent or correlated. The distribution of the set of these sample covariances is developed in both cases, thereby giving a formal basis for a diagnostic procedure that has been used to identify sources of negative variance component estimates and to reveal model deficiencies. This mixed or random analog of residual analysis, complete with diagnostic tools, is presented. This involves, in part, a re-examination of the model for mixed or random effects. The distribution applies to any random or mixed model and is illustrated here in actual repeated measures experiments and validated by simulations.

1.1 Introduction

The problem of estimating variance components is the equivalent of the problem of estimating the covariance, θ_t, between appropriately related observations. As alluded to in Hocking (1989), the estimate is an average of sample covariances, individually referred to herein as diagnostics, or is a simple linear function of such averages. Therefore, the development of the distribution theory for the variance component diagnostics will focus on the development of the distribution of the sample covariances. It will be useful to consider these as bilinear forms. For example, consider a three-factor factorial experiment with factor 1 random and factors 2 and 3 fixed. To estimate $\theta_1 = \phi_1$, a sample covariance of the form

$$C = 1/(a_1-1) \times \sum_{1}^{a_1}(\bar{Y}ijk. - \bar{Y}.jk.)(\bar{Y}ij^*k^*. - \bar{Y}.j^*k^*.)$$

is used, inwhich $j \neq j^*$ and $k \neq k^*$. This sample covariance can be written written as a bilinear form

$1/(n)(Z_1'AZ_2)$, with $Z_1' = (\bar{Y}ijk.)_1$, $Z_2' = \bar{Y}(ij^*k^*.)_1$,

$A = I_{a_1} - J_{a_1}J_{a_1}'/a_1$ and $n = a_1 - 1$.

Equivalently, the bilinear form can be written

$$1/(2n)\ [Z_1', Z_2']\ \begin{bmatrix} 0 & A \\ A & 0 \end{bmatrix}\ \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}. \qquad (1.1)$$

Except for rearranging indices, the bilinear form associated with any diagnostic can be written as (1.1). For simplicity, the examples discussed assume a three-factor model, but the methodology is general.

If a nonfactorial model is assumed, still with only one random factor and it is not nested, then a sample covariance of the form C is still appropriate. However, depending on the nesting, one of the conditions $j \neq j^*$, $k \neq k^*$ might be relaxed. In the case of four or more factors, the same results hold, so long as there is only one random factor (other than replication) and it is not nested.

The distribution of Z'_1AZ_2 depends on the covariance structure of (Z_1', Z_2'). There are two cases to consider. If there is only one random factor, such as factor 1, then $(Z_1', Z_2') \sim N(\mu, V)$, inwhich $\mu' = (\mu_1', \mu_2')$, and

$$V = \begin{Bmatrix} aI & cI \\ cI & aI \end{Bmatrix}$$

with each of a and c being a simple linear function of the variance components.

If factor 1 is not the only random factor, V may be more complex and the diagnostics are non-independent paired observations. This case will be discussed in section two.

1.2 Background

The first explicit density function for the bilinear form,

$$C = \sum_{i=1}^{a_1} (\bar{Y}ijk. - \bar{Y}.jk.)(\bar{Y}ij^*k^*. - \bar{Y}.j^*k^*.)/n,$$

with $n = (a_1-1)$ was developed by Pearson, Jeffery, and Elderton in 1929 based on independent sample pairs $(\bar{Y}ijk., \bar{Y}ij^*k^*.)$ having a bivariate normal distribution with the variance-covariance structure of V^* below. In summary, they used the result that if $\bar{Y}ijk.$ and $\bar{Y}ij^*k^*.$ are jointly normally distributed random variables, with expected values μ_1 and μ_2, respectively, and covariance

$$V^* = \begin{pmatrix} a & c \\ c & a \end{pmatrix},$$

then the conditional distribution of $\bar{Y}ijk.$, given $\bar{Y}ij^*k^*.$, is normal with expected value $(\mu_1+\rho)\ (\bar{Y}ij^*k^*. - \mu_2)$ and standard deviation $a(1-\rho^2)^{1/2}$. Thus, the conditional distribution of C, given the a_1 vector $(\bar{Y}ij^*k^*.)$ is normal with expected value $(\rho)S$ and standard deviation $(a(1-\rho^2)S)^{1/2}$,

where

$$S = \sum_{i=1}^{a_1} (\bar{Y}ij^*k^*. - \bar{Y}.j^*k^*.)^2.$$

As S is distributed $(a)\chi^2_{(a_1-1)}$, the probability density function of C is

$$f(c) = \frac{n}{(2\pi)^{1/2}a^{1/2}(1-\rho^2)^{1/2}} \int_0^\infty \frac{S^{(n-4)/2}\exp(-S/(2a))}{(2a)^{(n-3)/2}((n-1)/2)!} \times$$

$$\exp\left[-\frac{(nc-(\rho a^{1/2}S)/a)^2}{2a(1-\rho^2)S} \right] dS. \qquad (1.2)$$

Various other methods of deriving the distribution have been demonstrated by Wishart and Bartlett (1932), Hirschfeld (1937), and Mahalanobis, Bose, and Roy (1937).

Press (1967) presented some other equivalent forms of the density (1.3). Defining a sample of N independent observations $(Z_{11}, Z_{12}),\ldots, (Z_{N1}, Z_{N2})$ from $N_2(\mu, V^*)$, he found that for $n = N-1$,

$$f(c) = \frac{n\,(\beta^2 - \tau^2)^{n/2}(nc)^{(n-1)/2}e^{-\tau nc}}{\pi^{1/2}(2\beta)^{(n-1)/2}\Gamma(n/2)} K_{(n-1)/2}(|nc|), \qquad (1.3)$$

inwhich $K_\alpha(z)$ denotes a modified Bessel function, $\beta = \gamma\,\eta$, where γ and η are functions of ρ and the common variance (a) of the Z's, and are equal to $\gamma = [a(1-\rho_2)]^{-1}, \eta = [a(1-\rho_2)]^{-1}$, $\tau = \rho\beta$, and $\rho = c/a$ inwhich c is the covariance of the Z's and a the variance. When α is an integer, the Bessel function is referred to as a modified Bessel function, and when α is an odd half-integer, it is referred to as a modified spherical Bessel function of fractional order or a Bessel function of the third kind. When the number of degrees of freedom n is even, it is possible to express the density of C in terms of elementary functions and to calculate the exact expression since

$$K_{(n-2)/2}(z) = (\pi/(2z))^{1/2}e^{-z}\sum_{j=0}^{(n-3)/2}\frac{\Gamma((n-5)/2 + j)}{\Gamma(j+1)\ \Gamma((n-5)/2-j)\ (2z)^j}$$

Press (1967) provided formulae for computing the exact cumulative distribution function of the sample covariance for an even

number of degrees of freedom. In addition, percentage points of the C distribution for seven values of n and $\rho = 0$ were tabulated. However, for an arbitrary sample size, Press states that the probability density function of C "is a complicated expression which is difficult to evaluate." To evaluate the probability density function, it was necessary to develop an efficient formula for calculating the distribution function of the covariance utilizing the recursive properties of the Bessel function.

1.3 Distribution of the Sample Covariance for all Sample Sizes

In developing the computational formula of the distribution, two cases had to be considered. For the first case, N is even, and (N-2)/2 is an integer. The second case is that N is odd. Thus, the calculation of the probability density function requires calculation of the modified Bessel function for both integer and fractional order.

The computation of the modified Bessel function of integer order requires two polynomial approximations for order 0 and 1, which will be referred to in this paper as $k_0(y)$ and $k_1(y)$, respectively. These approximations are precise to at least $1x10^{-8}$. The approximations are defined in Abramowitz and Stegun (1964). From $k_0(y)$ and $k_1(y)$ and results in Abramowitz and Stegun (1964), the following recursive formula may be developed:

$$k_{(n+1)}(y) = (2n/y)\ k_n(y) + {}_{(n-1)}(y). \qquad (1.4)$$

For example, $k_2(y) = (2/y)\ k_1(y) + k_0(y)$,

and $k_3(y) = \frac{(4)k_2(y)}{y} + k_1(y)$.

The above formula (1.4) is useful in calculating the values of the (n-1)/2 order Bessel function. To determine the value of the Bessel function for fractional order the following relationship found in Abramowitz and Stegun (1964) was used:

$$\left(\frac{\pi}{2y}\right) K_{\left(\frac{n-1}{2}\right)}(y) = \frac{\pi}{2y} e^{-y} \sum_{k=0}^{m-1} \left\{ \frac{(m+k-1)!}{k!(m-k-1)!(2y)^k} \right\}$$

Given the values of the Bessel function for a fixed n, the probability density function of the distribution (1.3) was easily evaluated.

1.4 Calculation of CDF

The cumulative distribution function was computed using Simpson's integration method. Simpson's method of numerical integration approximates the probability density function by a set of parabolas. In general, Simpson's rule gives

$$\int_a^b f(x)dx = \left(f_0 + f_n + 4 \sum_{\substack{j=1 \\ \text{odd}}}^{n-1} f_j + 2 \sum_{\substack{j=2 \\ \text{even}}}^{n-2} f_j \right) \frac{\Delta x}{3},$$

where $\Delta x = (b-a)/n$, $f_j = f(a+j\Delta x)$.

1.5 Tabulated Cumulative Distribution for the Diagnostics

Critical percentile points of the covariance distribution for ρ ranging between -0.9 to 0.9 in increments of 0.1, with the sample size N between 2 to 10, 15, 20, 25, 30, 40 and 50, and the variances equal to one are contained in Grynovicki (1989). Specifically, this paper gives the value of C_{crit} such that $P[C \leq C_{crit}] = \alpha$, for $\alpha = 0.01$, 0.05, 0.10, 0.90, 0.95, and 0.99 inwhich C is the sample covariance from a bivariate normal with mean 0 and indicated variance-covariance matrix V.

1.6 CDF Program for Diagnostics

A computer program to calculate the cumulative distribution of the sample covariance (C/(N-1)) or equivalently the variance component diagnostics is presented in Grynovicki (1989). The program is written in Turbo-Pascal Version 4.0®, see Miller (1987), and can be compiled and run on any IBM-compatible or MacIntosh personal computer provided Turbo-Pascal 4.0 is available. The program utilizes Simpson's integration method and calculates the cdf using a tolerance of 10^{-6}.

1.7 Validation of Distribution

For ρ = -0.9 to 0.9, in increments of 0.1 and for sample size N = 2 to 10, and 50, a random sample of 1,000 sample covariances from a bivariate normal were generated as follow. First, three sets of 1,000 independent standard normal variates (Y_1, Y_2, Y_3) were generated using the Box-Muller transform. Second, 1,000xN independent samples from a bivariate normal distribution were generated with specified variances and a covariance using the transformation

$$Z^*_1 = \sigma_1 \left(\sin (A_1)\ Y_1 + \cos (A_2)\ Y_2 \right) \text{ and}$$

$$Z^*_2 = \sigma_2 \left(\sin (A_2)\ Y_3 + \cos (A_2)\ Y_2 \right)$$

in which

$$A_1 = \arccos[(|\sigma_{12}|/(\sigma_1\sigma_2))^{1/2}], \text{ and}$$

$$A_2 = A_1 \qquad \text{if } \sigma_{12} \geq 0,$$

$$= \pi - A_1 \qquad \text{if } \sigma_{12} < 0.$$

Finally, the 1,000 covariances were calculated by sequentially selecting 1,000 pairs (Z^*_1, Z^*_2) of N-vectors and calculating the covariance $Z^*_1{}'AZ^*_2$, where A = I-JJ'/N, I is NxN identity matrix, and J is a N column vector of 1's.

As a partial check of the density function, a comparison of the simulation and actual distribution was made using the Kolmogorov-Smirnov one-sample goodness-of-fit test. The test statistic is

$$D = \text{maximum } |F(x) - S(x)|, \ -\infty < x < \infty,$$

inwhich F and S are the theoretical and simulated distribution functions, respectively. For a sample size of 1,000, the critical value of this statistic is 0.043 at $\alpha = 0.05$.

Comparison of the theoretical and simulated values was made for values of N from 2 to 10 and 50 for values of ρ in increments of 0.1, between -0.9 and 0.9, and for variances equal to one. Two SAS computer programs were written to generate the simulated value and to calculate the Kolmogorov-Smirnov maximum deviation statistic. These programs are contained in Grynovicki (1989).

The calculated D for the specified parameters can be found in Grynovicki (1989). All 190 simulations were determined to have a calculate D below 0.043. Thus, the simulated distribution is consistent with the one derived when compared at the 0.05 probability level.

It is worth noting that the maximum deviations occurred at the center of the distribution and not at the tails.

1.8 Validation of Distribution for Diagnostic Tables

1.8.1 Introduction

Once the distribution for independent diagnostics was developed and validated, the next step was to determine if the distribution could be used in evaluating a table of diagnostics that are correlated. Searle (1971b) has shown that the correlation of two bilinear forms is equal to

$$\text{Cov}(Z'_1A_{12}Z_2,\ Z'_3A_{34}Z_4) = \text{tr}(A_{12}C_{23}A_{34}C_{41} + A_{12}C_{24}A_{34}C_{31}),$$

in which $E(Z_1) = E(Z_2) = E(Z_3) = E(Z_4) = 0$,

where $C_{uv} = \text{Cov}(Z_u, Z_v)$, if $u \neq v$ and

$= \text{Var}(Z_u, Z_v)$, if $u = v$.

Also define $Z' = [Z_1', Z_2', Z_3', Z_4']$, so that $Z \sim N(0,V)$, inwhich

$$V = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{pmatrix}.$$

To determine how well the derived distribution fits correlated diagnostics, an experiment will be simulated at least 200 times and the calculated diagnostics will be compared with the theoretical distribution. For simplicity, I will consider a 3-way factorial experiment with factor 1 random and factors 2 and 3 fixed. In this simulation, θ_1, the covariances of the form

$$C_1 = 1/(a_1-1)\sum_i(\bar{Y}ij_1k_1. - \bar{Y}.j_1k_1.)(\bar{Y}ij_2k_2. - \bar{Y}.j_2k_2.) = 1/(a_1-1)\ Z_1'A_{12}Z_2,$$

in which $j_1 \neq j_2$ and $k_1 \neq k_2$ will be the diagnostics used. Also define

$$C_2 = 1/(a_1-1)\sum_i(\bar{Y}ij_3k_3. - \bar{Y}.j_3k_3.)(\bar{Y}ij_4k_4. - \bar{Y}.j_4k_4.) = 1/(a_1-1)\ Z_3'A_{34}Z_4.$$

For this experiment if we let

card $j = \#(j_1, j_2) \cap (j_3, j_4)$,

card $k = \#(k_1, k_2) \cap (k_3, k_4)$, and

card $jk = \#((j_1, k_1), (j_2, k_2)) \cap ((j_3, k_3), (j_4, k_4))$,

then the covariance of any two of the diagnostics for θ_1 is

$\text{cov}(C_1, C_2) = 2\theta_1^2$ if card j = card k = 0,

$\theta_1^2 + \theta_1\theta_{12}$ if card j = 1, card k = 0,

$\theta_1^2 + \theta_1\theta_{13}$ if card j = 0, card k = 1,

$\theta_1^2 + \theta_1\theta_{123}$ if card j = card k = card jk = 1,

$\theta_1\theta_{12} + \theta_1\theta_{13}$ if card j = card k = 1, card jk = 0,

$\theta_1^2 + \theta_{13}\theta_{123}$ if card j = 1, card k = 2, card jk = 1,

$\theta_1^2 + \theta_{12}\theta_{123}$ if card j = 2, card k = card jk = 1.

Also, the var $(C_1) = \theta_1^2 + \theta_{123}^2$.

Other experimental designs are entirely analogous. If a nonfactorial model is assumed with only one non-nested random factor, a sample covariance of the form C is still appropriate although, depending on the nesting, one of the conditions, $j_1 \neq j_2^*$, $k_1 \neq k_2^*$ might be relaxed. The variance-covariance matrix V

still has the form assumed even though the variance and covariance may be different functions of the variance components.

1.8.2 Simulation of a Three-Factor Factorial Experiment

The linear model used in this simulation was

$$Yijkt = M + Ai + ABij + ACik + ABCijk + \epsilon(ijkt).$$

Here, M represents the grand mean and all fixed effects, and the remaining terms are independent distributed normal with mean zero and variance given by the associated variance component. The structure of the covariance matrix for this design as defined in Hocking (1985) is

$$V = \lambda_1(A_0 + A_1) + \lambda_{12}(A_2 + A_{12}) + \lambda_{13}(A_3 + A_{13}) + \lambda_{23}(A_{23} + A_{123}), \qquad (1.5)$$

where $A_i = (1/a^*_i)\ G_1 \otimes G_2 \otimes \ldots G_k \otimes J_{a_1}J_{a_1}'$, $a^*_i = \Pi_{i \in a_i}$,

$G_i = I_{a_i} - 1/a_i J_{a_i} J'_{a_i}$ if $i \notin T$ or $J_{a_i}J_{a_i}$ if $i \in T$,

and λ_i are the eigenvalues of V.

For this model, the variance for Z_i associated with the terms comprising the bilinear form has variance $Var(Z_i) = \phi_1 + \phi_{12} + \phi_{13} + \phi_{123}$. Its covariance is $cov(Z_iZ_j) = \phi_1$.

Two cases of this design were considered. For the first case $a_1 = 3$, $a_2 = 3$ and $a_3 = 2$. In the second case, $a_1 = 3$, $a_2 = 3$, and $a_3 = 4$. In the first case, 500 $a_1 \times a_2 \times a_3$ independent sample from a standard normal distribution were generated and in the second

case, 200 $a_1 \times a_2 \times a_3$ were generated. Both used Box-Muller. Then, a sample of size $a_1 \times a_2 \times a_3$ was sequentially selected and multiplied by $V^{1/2}$ where $V^{1/2}$ is the same as formula 1.5 except that the eigenvalues are replaced by its square root.

For case one, 6 diagnostics were generated per iteration and in the second case 36 diagnostics were generated giving 3,000 diagnostics for case one and 7,200 diagnostics for case two. The value of the variance components was varied to obtain values of ρ between -0.4 and 0.8. Due to the positive definiteness of the variance covariance matrix V, -0.4 was the smallest value one could expect from this design. The results for both cases are shown in Table 1.1. For case one, the maximum difference for the simulation and theoretical distribution ranged between 0.037 and 0.11. However, for the critical probabilities of .01, .05, and .1, the estimated critical values were small and conservative. The $P(C \leq C_{crit})$ was always larger than what the simulation showed. The difference in the agreement between the theoretical and simulation increased as one increased in probability from 0.01 to 0.10. The maximum difference in the two distributions occurred in the center of the distribution. For the high critical values in case one and all critical values in case two, the simulation and theoretical distribution agreed. The maximum deviation between the theoretical and simulation ranged between .009 and .017 for case two. As in case one, the estimated critical values were conservative.

TABLE 1.1

Calculated Kolmogorov-Smirnov One Sample Statistic, D, and Probability Differences at Critical Values for Simulated and Theoretical Distribution of Variance Component Diagnostics for Various Values of ρ when Variances are Equal

$a_2 = 3 \quad a_3 = 2$

		Difference at Critical Probabilities α					
ρ	D	.01	.05	.10	.90	.95	.99
-0.43	0.094	0.007	0.036	0.059	0.027	0.021	0.009
-0.21	0.110	0.008	0.040	0.071	0.016	0.016	0.007
-0.09	0.079	0.009	0.035	0.069	0.011	0.010	0.005
0.04	0.081	0.007	0.038	0.062	0.002	0.007	0.004
0.15	0.096	0.009	0.039	0.071	0.009	0.004	0.003
0.25	0.061	0.001	0.038	0.057	0.000	0.006	0.000
0.41	0.064	0.009	0.036	0.055	0.007	0.004	0.002
0.61	0.072	0.000	0.033	0.058	0.006	0.002	0.001
0.82	0.037	0.009	0.029	0.022	0.002	0.009	0.001
			$a_2 = 3 \quad a_3 = 4$				
-0.43	0.017	0.001	0.003	0.002	0.007	0.003	0.002
-0.21	0.012	0.000	0.001	0.000	0.009	0.001	0.001
-0.09	0.013	0.001	0.001	0.002	0.008	0.002	0.001
0.04	0.015	0.002	0.001	0.004	0.008	0.003	0.002
0.15	0.021	0.000	0.000	0.001	0.009	0.005	0.000
0.25	0.013	0.002	0.001	0.003	0.006	0.003	0.001
0.41	0.010	0.003	0.002	0.005	0.008	0.003	0.000
0.61	0.009	0.002	0.003	0.007	0.006	0.004	0.001
0.82	0.014	0.002	0.008	0.011	0.005	0.006	0.006

Based on these findings, one can use the table of diagnostics to identify abnormally large or small covariances in the table. This diagnostic method will allow researchers the tool to investigate sources of negative variance component estimates, identify outliers and reveal model deficiencies.

Having developed the distribution of the diagnostics for bilinear form when the sample is from a set of independent observations distributed $N_2(\mu, V)$, the next step is to develop the distribution for the diagnostics (covariance) in which the assumption of independent paired observations does not hold. The development of this distribution and its validation is presented below.

2.1 Distribution Theory for the Variance Component Diagnostic for Non-Independent Paired Observations

The final phase in developing the distribution theory for the variance components was to consider the case where the sample pairs (Z_{1j}, Z_{2j});$(j = 1, 2, \ldots, a_2)$; are from a bivariate normal distribution with variance-covariance structure

$$V = \begin{pmatrix} aI + bJJ' & cI + dJJ' \\ cI + dJJ' & aI + bJJ' \end{pmatrix}$$

The small letters represent linear combinations of the variance components as specified by the linear model, I is an identity matrix, and J is a column of ones. This circumstance arises when dealing with a linear model of more than one random main effect and then only in regard

to certain variance components associated with the interaction.

The Representation Theorem presented in Green (1987) allows the diagnostics for designs of all sizes to be estimated in an unbiased and efficient manner, regardless of the number of random factors, or type of nesting. This theorem states that complex diagnostics can be written as a linear combination of simpler sample covariances. Each sample covariance is based on the levels of a single factor. Thus, the only bilinear forms required are of the type $Z_1'AZ_2$, in which Z_1 and Z_2 are vectors of responses that vary the levels of only one factor, and $A=I-JJ'/a_i$, in which a_i is the number of levels of that one factor. Thus, in developing the distribution of the diagnostics for paired samples which are not independent, and having already attained the distribution for the independent case, the distribution of the diagnostics for any design with at least one random factor will be completed.

2.2 Helmert Transformation

The first step in developing this distribution was to determine a transformation that could change the variance covariance structure so that the transformed paired observations would be independent and have the variance-covariance structure

$$V = \begin{pmatrix} aI & cI \\ cI & aI \end{pmatrix}$$

Using all but the first row of the Helmert matrix as the matrix of the transformation, it will be shown that the bilinear form

$$Z_1'AZ_2 = X_1'X_2 = X_1'A^*X_2 + (a_2-1)\bar{X}_1\bar{X}_2. \qquad (2.1)$$

in which $X_1 = WZ_1$, W is the Helmert matrix excluding the first row,

$A = I_{a_2}-(J_{a_2})(J_{a_2}')$, and $A^* = \left(I_{a_2-1} -(J_{a_2-1})(J_{a_2-1}')\right)$

The Helmert matrix, H, is an orthonormal matrix. The first row of H is $J'/(a_2)^{1/2}$. For $r = 2, \ldots, a_2$, the r^{th} row of H has its first r-1 components equal to $[r(r-1)]^{-1/2}$, the r^{th} component equal to $-(r-1)/[r(r-1)]^{1/2}$, and the remaining components equal to 0.

PROOF OF 2.1:

Let $Z_1' = (\bar{Y}i1k., \bar{Y}i2k., \ldots, \bar{Y}ia_2k.)$,

$Z_2' = (\bar{Y}i1k^*., \bar{Y}i2k^*., \ldots, \bar{Y}\ ia_2k^*.)$,

$Z_1 \approx N_{a_2}(0, C11)$, and

$Z_2 \approx N_{a_2}(0, C22)$, with

$C11 = aI + bJJ'$,

$C22 = aI + bJJ'$, and

$Cov(Z_1,Z_2) = C12 = cI + dJJ'$, in which a, b, c, and d are linear functions of the variance components.

The Helmert matrix H is:

$$H' = \left[a_2^{-1/2}J,\ W_2'\right] = [W_1',\ W_2'].\quad \text{Also,}$$

$$X_1 = HZ_1 = \begin{bmatrix} W_1Z_1 \\ W_2Z_1 \end{bmatrix} = \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix} \text{ and}$$

$$X_2 = HZ_2 = \begin{bmatrix} W_1Z_2 \\ W_2Z_2 \end{bmatrix} = \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix}.$$

Then, $X_1'X_2 = Z_1'H'HZ_2 = Z_1'Z_2$, since H is orthonormal, and

$X_1'X_2 = Z_1'W_1'W_1Z_2 + Z_1'W_2'W_2Z_2 = a_2\bar{Z}_1\bar{Z}_2 + Z_1'W_2'W_2Z_2$.

Rearranging terms,

$$\begin{aligned} X_{12}'X_{22} &= Z_1'W_2'W_2Z_2 \\ &= X_1'X_2 - a_2\bar{Z}_1\bar{Z}_2 \\ &= Z_1'Z_2 - a_2\bar{Z}_1\bar{Z}_2 \\ &= Z_1'AZ_2. \end{aligned}$$

Since $A = (I - JJ'/a_2)$ it follows that $X_{12}'X_{22} = X_{12}'AX_{22} + (a_2-1)\bar{X}_{12}\bar{X}_{22}$. Thus, the bilinear form $Z_1'AZ_2$ is equal to $X_{12}'X_{22}$. Now, the variance covariance structure of (X_{12}, X_{22}) is of the form

$$V = \begin{pmatrix} aI & cI \\ cI & aI \end{pmatrix},$$

since $W_2C_{11}W_2' = aI$, $W_2C_{22}W_2' = aI$, and $W_2C_{12}W_2' = cI$. Having established that the bilinear form $Z_1'AZ_2 = X_{12}'AX_{22} + (a_2-1)\bar{X}_{12}\bar{X}_{22}$, the next step was to determine the distribution of $X_{12}'AX_{22} + (a_2-1)\bar{X}_{12}\bar{X}_{22}$.

2.3 Distribution of Transformed Variables

First, one must realize that the bilinear form can be written as a linear combination of central chi-squares and that $(a_2-1)\,\bar{X}_{12}\bar{X}_{22}$ can be written as a linear combination of chi-squares.

Specifically, a property of the bilinear form is that

$$X_{12}'A_{a_2-1}X_{22} = a\frac{\left[(1+\rho)\,\chi^2_{(a_2-2)}-(1-\rho)\chi^2_{(a_2-2)}\right]}{2},$$

in which a is the common variance of X_{12} and X_{22}, ρ is the correlation between X_{12} and X_{22}, and $\chi^2_{(a_2-2)}$ is the central chi-square with (a_2-2) degrees of freedom.

PROOF:

Consider the product, X_1X_2, of deviations from the sample mean inwhich X_1 and X_2 are singletons.

Let $X' = (X_1, X_2)$. Then $X = (X_1, X_2)' \approx N_2(0, V^*)$, where

$$V^* = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

If $A = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$, then $X_1X_2 = (X_1,X_2)\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}(X_1,X_2)'$.

The characteristic function is given by $E(e^{itX'AX})$

$$= \int\int \frac{1}{2\pi|V|^{1/2}} \exp\left[itX'AX - 1/2\,(X_1,X_2)\,V'\,(X_1,X_2)'\right] dX_1\, dX_2,$$

which, since $\left((I-2itAV)V^{-1}\right)^{-1} = V(I-2tiAV)^{-1}$, may be written as

$$\iint \frac{1}{2\pi|V|^{1/2}} \exp\left[-\frac{1}{2}(X_1,X_2)\left((I-2itAV)V^{-1}\right)^{-1}(X_1,X_2)'\right] dX_1\, dX_2.$$

Let $W = \left[(I-2itAV)V^{-1}\right]^{-1}$.

By the identity $2\pi^{n/2}|W|^{1/2} = \int\ldots\int \exp\left[-1/2X'W^{-1}X\right] dX_1\ldots dX_n$, one obtains,

$$= V^{1/2}\left(|V(1-2itAV)^{-1}|\right)^{1/2}$$

$$= (|V-2itAV|)^{-1/2}$$

$$= 1/2\begin{vmatrix} 1-ti\rho\sigma_1\sigma_2 & -ti\sigma_2^2 \\ -ti\sigma_2^2 & 1-ti\rho\sigma_1\sigma_2 \end{vmatrix}^{1/2}$$

$$= \left(1-2it\rho\sigma_1\sigma_2 + t^2(1-\rho^2)\sigma_1^2\sigma_2^2\right)^{1/2}$$

$$= \left[\left(1-\frac{2it\sigma_1\sigma_2(1+\rho)}{2}\right)\left(1+\frac{2it\sigma_1\sigma_2(1-\rho)}{2}\right)\right]^{-1/2}.$$

It follows that $X'AX = \frac{\sigma_1\sigma_2}{2}\left[(1+\rho)K_1 - (1-\rho)K_2\right]$ in which K_1 and K_2 are independent χ_1^2. If $Z_1^* = (Y_{11}, Y_{12},\ldots, Y_{1a_2-1})$ and $Z_2^* = (Y_{21}, Y_{22},\ldots, Y_{2a_2-1})$ then

$$Z_1^{*\prime}AZ_2^* = X_1'X_2 = X_1'IX_2,$$

where $X_1 = WZ_1^*$, W is the a_2-2 rows of the Helmert matrix, and $A = I_{a_2-2} - J_{a_2-2}J'_{a_2-2}$. Then, the characteristic function of $Z_1^*AZ_2^*$ is

$$E\left(\exp^{itZ_1^*AZ_2^*}\right) = E\left(\exp^{itX_1'X_2}\right)$$

$$= \prod_{1}^{a_2-2}\left[\left(1-\frac{2it\sigma_1\sigma_2(1+\rho)}{2}\right)\left(1+\frac{2it\sigma_1\sigma_2(1-\rho)}{2}\right)\right]^{-1/2}.$$

Thus, the distribution of $Z_1^*AZ_2^*$ is equivalent to the distribution of $1/2\sigma_1\sigma_2[(1+\rho)K_1-(1-\rho)K_2]$, where K_1 and K_2 are independent chi-square variables with a_2-2 degrees of freedom.

Second, one must show that $(a_2-1)\bar{X}_1\bar{X}_2$ is distributed as a linear combination of central chi-squares. Specifically, if one defines $\bar{Y}_1 = (a_2-1)^{1/2}\bar{X}_1$ and $\bar{Y}_2 = (a_2-1)^{1/2}\bar{X}_2$, then $(\bar{Y}_1, \bar{Y}_2) \approx N_2(0,Z)$, where

$$Z = \begin{bmatrix} a & c \\ c & a \end{bmatrix},$$

and the distribution of $(a_2-1)\bar{X}_1\bar{X}_2$ is also that of $(a+c)\chi_1^2-(a-c)\chi_1^2$.

PROOF:

Let $\bar{X} = (\bar{X}_1,\bar{X}_2)' = \frac{1}{(a_2-1)}\begin{bmatrix} J' & 0 \\ 0 & J' \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$,

in which $X = WZ$, as previously defined. Then

$X \approx N_{a_2-1}(0,V)$, and $\bar{X} \approx N_2(0,1/(a_2-1)Z)$, where

$$1/(a_2-1)Z = \frac{1}{(a_2-1)^2}\begin{bmatrix} J' & 0 \\ 0 & J' \end{bmatrix}\begin{bmatrix} aI_{a_2-1} & cI_{a_2-1} \\ cI_{a_2-1} & aI_{a_2-1} \end{bmatrix}\begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix}$$

$$= \frac{1}{a_2-1}\begin{bmatrix} a & c \\ c & a \end{bmatrix}.$$

Let $\tilde{Y}_1 = (a_2-1)^{1/2}\tilde{X}_1$. Then $\tilde{Y} \approx N_2(0,Z)$ and $\tilde{Y}_1\tilde{Y}^2 = (a_2-1)\tilde{X}_1\tilde{X}_2$.

Define a 2X2 Helmert matrix, $H = \frac{1}{(2)^{1/2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then

$W' = (W_1', W_2') = (H\tilde{Y})' = \left[(\tilde{Y}_1+\tilde{Y}_2)/(2)^{1/2}, (\tilde{Y}_1-\tilde{Y}_2)/(2)^{1/2} \right]$, and

$$W \approx N_2\left(0, \begin{pmatrix} a+c & 0 \\ 0 & a-c \end{pmatrix} \right)$$

Thus $(\tilde{Y}_1+\tilde{Y}_2)/(2)^{1/2}$ and $(Y_1-\tilde{Y}_2)/(2)^{1/2}$ are independently normally distributed with variance (a+c) and (a-c), respectively. By Theorem 2.3 in Hocking (1984),

$$W_1'W_1 \approx (a+c)\chi^2_1,$$

$$W_2'W_2 \approx (a-c)\chi^2_1, \text{ and}$$

$$W_1'W_1 + W_2'W_2 = \tilde{Y}_1\tilde{Y}_2 = (a_2-1)\tilde{X}_1\tilde{X}_2.$$

2.4 Distribution of Linear Combinations of Weighted Central Chi-Squares

Define $C = T_1 - T_2$, in which

$$T_1 = a_1 \left[\chi^2_{a_2-2} + \frac{a+c}{a_1} \chi^2_1 \right], \text{ and}$$

$$T_2 = b_1 \left[\chi^2_{a_2-2} + \frac{a-c}{b_1} \chi^2_1 \right],$$

$$a_1 = \frac{a}{2}[(1+\rho)],$$

$$b_1 = \frac{a}{2}[(1-\rho)],$$

$a_1 > 0$, $b_1 > 0$, $a^* = (a+c)/a_1 > 1$, $b^* = (a-c)/b_1 > 1$, and all the chi-squared variates are independent. The distribution of T_1 can be represented by

$$F_{T_1}(x) = \sum_{i=0}^{\infty} q_i \, F_{(a_2-1)+2i}(x/a_1),$$

inwhich $\sum_{i=0}^{\infty} q_i = 1$ and the q_i are weight constants depending on $(a+c)/a_1$ and a_2. The weight constant q_r is equal to

$$q_r = \frac{a_1^{-1/2}(1-(1/a_1))^r \Gamma(r+1/2)}{\Gamma(r+1)\ \Gamma(1/2)} \text{ for } r \geq 0,$$

in which $\Gamma(1/2) = \pi^{1/2}$, and

$$\Gamma(r + 1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2r-1)\sqrt{\pi}}{2^r}.$$

PROOF:

Let $\phi_{\chi^2_n}(t)$ denote the characteristic function of a central chi-square with n degrees of freedom and $\Psi_{T_1}(t)$ the characteristic

function of T_1. Then, $\phi_{\chi^2_n}(t) = (1-2it)^{-n/2}$, and, because the chi-squared variates are independent,

$$\Psi_{T_1/a_1}(t) = 2\phi_{\chi^2_{(a_2-2)}}(t)\phi_{a^*\chi^2_1}(t).$$

The characteristic function of a constant times a central chi-squared variate is given by Robbins, Herbert, and Pitman (1949) as

$$\begin{aligned}\phi_{a^*\chi^2_n}(t) &= (1-2ia^*t)^{-n/2} = [a^*(1-2it)-(a^*-1)]^{-n/2} \\ &= a^{*-n/2}(1-2it)^{-n/2}\left[1-(1-1/a^*)(1-2it)^{-1}\right] \\ &= a^{*-n/2}(1-2it)^{-n/2}\left[1-(1-1/a^*)(1-2it)^{-1}\right]^{-n/2}. \qquad (2.2)\end{aligned}$$

By the binomial theorem, we have for $a^* > 0$,

$$a^{*-n/2}[1-(1-1/a^*)Z]^{-n/2} = \sum_j q_j Z_j \text{ for } |Z| < |1-1/a^*|^{-1}, \qquad (2.3)$$

$a^* \geq 1$, $q_j \geq 0$ $(j = 0, 1, \ldots)$, and $\sum q_j = 1$. Since $|1-2it|^{-1} \leq 1$, for all real t it follows from (4.3) that for $a^* \geq 1$,

$$(1-2ia^*t)^{-n/2} = \sum q_j (1-2it)^{-n/2-j} = \sum_j q_j \phi_{\chi^2_{n+2j}}(t).$$

Now, the characteristic function of T_1/a_1 may be obtained from (4.2) and the following defining identity for the constants q_j's, where $N = a_2-1$.

$$\left[\, a^{*-N/2}\left[1-(1-1/a^*)Z\right]^{N/2}\,\right] = \sum q_j Z^j, \quad (|Z| \leq 1).$$

It follows that

$$\Psi_{(T/a_1)}(t) = (1-2it)^{-(N/2)}\left[a^{*-1/2}\left[1-(1-1/a^*)(1-2it)^{-1}\right]^{-1/2}\right],$$

$$= \sum_{j=0}^{\infty} q_j (1-2it)^{-(N/2+j)},$$

$$= \sum_{j=0}^{\infty} q_j \phi_{\chi^2_{(N+2j)}}(t),$$

which is the characteristic function of T_1/a_1. Hence, the cdf of T_1/a_1 by inversion is $\sum_{i=1}^{\infty} q_i \, F_{N+2i}(t)$, where $F_{N+2i}(t)$ denotes the cumulative distribution function of a central chi-square with N+2i degrees of freedom.

It follows on setting $X = a_1 T$ that the cdf of T_1 is given by

$$\sum_{i=1}^{\infty} q_i \, F_{N+2i}(x/a_1).$$

Similarly, the cdf of T_2 is given by

$$F_{T_2}(w) = \sum_{j=0}^{\infty} q_j F_{N+2j}(w/b_1).$$

Since T_1 and T_2 are linear combinations of central chi-squared variates, if f_{T_1} and f_{T_2} denote the densities of T_1 and T_2 respectively, then the pdfs of T_1 and T_2 are given by

$$f_{T_1} = \sum_i (q_i/a_1) \, f_{N+2i}(x/a_1), \text{ and}$$

$$f_{T_2} = \sum_j (q_j/b_1) \, f_{N+2j}(w/b_1).$$

PROOF:

$$\sum q_i F_{N+2i}(x/a_1) = \sum q_i \int_{-\infty}^{\infty} F_{N+2i}(x/a_1).$$

By the Beppo-Levy Theorem (Morrison, 1987), this

$$= \int_{-\infty}^{\infty} \sum_I q_i F_{N+2i}(x/a_1) \quad = F_{T_1/a}(x/a_1).$$

Now, by Fubini's theorem (Wheeden and Zygmund, 1977),

$$F' = \left(\sum_I q_i F_{N+2i}(x/a_1) \right)' = \sum_I q_i F'_{N+2i}(x/a_1) = \sum_I q_i f_{N+2i}(x/a_1).$$

2.5 Probability Density for Diagnostics

In this section, the probability density function for T_1-T_2 = C, which is the diagnostic when the sample pairs are not independent, will be developed. Let f(x) and g(w) denote the pdfs of T_1 and T_2 respectively. By convolution, the pdf of $C = T_1-T_2$ is

$$h(t) = \int_{-\infty}^{\infty} f(t+w)\ g(w)\ dw. \qquad (2.4)$$

In the previous section, we have shown that

$$f(x) = \sum_I (q_i/a_1)\ f_{N+2i}(x/a_1),\ x \geq 0, \text{ and} \qquad (2.5)$$

$$g(w) = \sum_J (q_j/b_1)\ f_{N+2j}(w/b_1),\ w \geq 0. \qquad (2.6)$$

Since the series converge uniformly, permitting interchange of integration and summation, we may substitute (4.5) and (4.6) into (4.4), and letting M = N, one obtains

$$h(t) = \sum_i \sum_j q_i q_j / (a_1 b_1) \int_0^{\infty} f_{M+2i}((t+w)/a_1)\ f_{M+2j}(w/b_1)\ dw =$$

$$\sum_i \sum_j \frac{q_i q_j\ t^{(2M+2i+2j\ -\ 2)/2} e^{-(t/2a_1)}}{2^{(M+2i+M+2j)/2} a_1^{(M+2i)/2} b_1^{(M+2j)/2} \Gamma((M+2i)/2)\ \Gamma((M+2j)/2)} \times$$

$$\left[\int_0^{\infty} e^{-((a_1+b_1)/(2a_1b_1))w}\ w^{(M+2j-2)/2}\ (1+w)^{(M+2i-2)/2} \right] dw. \qquad (2.7)$$

It is worth noting that the integral given below,

$$1/\Gamma(M+2j/2) \int_0^{\infty} e^{-((a_1+b_1)/(2a_1b_1))w}\ w^{(M+2j-2)/2}\ (1+w)^{(M+2i-2)/2}\ dw,$$

is the confluent hypergeometric function and is identical with the function U(a, b, x) discussed by Slater (1960). Having obtained the distribution of the diagnostic, the problem of how to evaluate it remained. This required the development of new recurrence relations for the definite integral.

2.6 Distribution of Bilinear Form from Non-Independent Bivariate Normal

It has been shown above that $Z_1'AZ_2 = X_1'X_2 + (a_2-1)\,\bar{X}_1\bar{X}_2$, in which X_i is the Helmert transformed data. If $a_2 = 2$, then $Z_1'AZ_2 = X_1'X_2 = (a_2-1)\,\bar{X}_1\bar{X}_2$. It has also been shown that $(a_2-1)\,\bar{X}_1\bar{X}_2 \approx (a+c)\chi^2_1 - (a-c)\chi^2_1$, where a is the variance of X and c is the covariance.

In the linear model context, the variance (a) can be broken down into a set of variance components comprising the

covariance (β), as well as a set that is not contained in the covariance (α). Therefore, defining the variance as $a = \alpha+\beta$ and the covaraince as $b = \beta$, the distribution of

$$Z_1'AZ_2 = X_1X_2 \approx \frac{2(\alpha+\beta)}{2}\left[\left(1+\frac{\beta}{\alpha+\beta}\right)\chi^2_1 - \left(1-\frac{\beta}{\alpha+\beta}\right)\chi^2_1\right].$$

Therefore, for $N = 2$, the distribution of the bilinear form is the distribution of the covariance from independent paired observations with twice the estimated variance.

2.7 Development of New Confluent Hypergeometric Recurrence Relations

2.7.1 Relation of Hypergeometric and Bessel Function

The calculation of the cdf for the bilinear form when the sample pairs are not independent required the development of new recurrence relations for the confluent hypergeometric function. In the notation of Abramowitz and Stegun (1964), equations 13.1.10 and 13.2.5, U(a, b, x) is the confluent hypergeometric function of Kummer and is given by

$$U(a,\ b,\ x) = 1/\Gamma(a)\ \int_0^\infty e^{-xt}\cdot t^{a-1}(1+t)^{b-a-1}\ dt.$$

Abramowitz and Stegun give two special cases for which U(a, b, x) can be written in terms of the modified Bessel functions. Using these relationships, initial values of the confluent hypergeometric function for the cdf were obtained as follows.

For the case N is odd and i = j, let r = (N - 1)/2 + i, and x = 2z. Then 2r + 1 = N + 2i = N + i + j and r + 1/2 = N/2 + i = N/2 + j. Using Abramowitz and Stegun equation 13.6.21,

$$U(N/2 + j, N + i + j, x) = U(r + 1/2, 2r + 1, 2z)$$

$$= \pi^{-1/2} e^{z} (2z)^{-r} K_{r}(z)$$

$$= \pi^{-1/2} e^{x/2} x^{-(N-1+2i)/2} K_{(N-1+2i)/2}(x/2).$$

For the case N is even and i = j, let r = (N - 2)/2 + i. Then r + 1 = N/2 + i and 2r + 2 = N + 2i = N + i + j. Using Abramowitz and Stegun equation 13.6.24,

$$U(N/2 + j, N + i + j, x) = U(r + 1, 2r + 2, 2z)$$

$$= \pi^{-1/2} e^{z} (2z)^{-(2r+1)/2} K_{(2r+1)/2}(z)$$

$$= \pi^{-1/2} e^{x/2} x^{-(N-1+2i)/2} K_{(N-1+2i)/2}(x/2).$$

Note that this expression is identical to the one obtained for odd N. Now by choosing i = j = 0 and i = j = 1 with a = N/2 and b = N one is now able to calculate two values for the confluent hypergeometric function for a given value of x. Specifically,

$$U(N/2, N, x) = U(a, b, x) \text{ and}$$

$$U(N/2 + 1, N + 2, x) = U(a + 1, b + 2, x).$$

From these two starting values, a recurrence relation is needed to obtain the remaining cases involved in calculating the probability density function.

2.7.2 New Recurrence Relations for Confluent Hypergeometric Functions

The evaluation of the pdf depended on being able to calculate U(a, b + 1, x) and U(a + 1, b + 1, x). From Abramowitz and Stegun equations 13.4.16, 13.4.18, and 13.4.19, replacing a with a + 1 and b with b + 1 in 13.4.16 and 13.4.18, one obtains

$$(a+x)\,U(a, b, x) - xU(a, b+1, x) + a(b-a-1)U(a+1, b, x) = 0, \quad (2.7.1)$$

$$(b-a-1)U(a, b-1, x) + (1-b-x)U(a, b, x) + xU(a, b+1, x) = 0, \quad (2.7.2)$$

and $$(b-a)U(a, b, x) + U(a-1, b, x) - xU(a, b+1, x) = 0. \quad (2.7.3)$$

From these, it follows that

$$(b - a)(b - a - 1)U(a + 1, b, x) + (b + x)U(a, b + 1, x) = x\,(a + x)U(a + 1, b + 2, x). \quad (2.7.4)$$

Now, 4.7.1 and 4.7.4 are two equations in the two unknowns U(a + 1, b, x) and U(a, b + 1, x) and the known quantities U(a, b, x) and U(a + 1, b + 2, x). The solutions by Cramer's rule are

$$U(a+1, b, x) = \frac{(x^2)\,U(a+1, b+2, x) - (b+x)\,U(a, b, x)}{b\,(b-a-1)}, \text{ and}$$

$$U(a, b+1, x) = \frac{(ax)U(a+1, b+2, x) + (b-a)\,U(a, b, x)}{b}.$$

From these, using recurrence relation 13.4.16 in Abramowitz and Stegun, with b replaced by b + 1, U(a, b + 2, x) can be calculated in

terms of U(a, b, x) and U(a, b + 1, x). The process can then be continued to calculate U(a, b + 3, x) and all other values of b for a specific a value. Similarly recurrence relation 13.4.17, with a replaced by a+1, gives starting values U(a + 1, b + 1, x) and U(a + 1, b + 2, x). Other entries are obtained for the remaining a+1 elements by using the same recurrence relation. These recurrence relations were used iteratively to calculate the U functions for fixed i and all j. Thus, the cdf can be evaluated.

2.8 Turbo Program for Diagnostics from Non-Independent Observations

A computer program to calculate the cumulative distribution of linear combinations of central chi-squared variables or equivalently, the variance component diagnostics based on non-independent paired observations are presented in Grynovicki and Green (1990). The program is written in Turbo-Pascal and can be compiled and run on any IBM compatible personal computer on which Turbo-Pascal is available. The program utilizes Simpson's integration method and calculates the cdf using a tolerance of 0.0000006.

2.9 Validation of the Distribution for the Diagnostics

For ρ between -0.2 to 0.8 the theoretical distribution was compared to the diagnostics for θ_{12} from a three-way hierarchical experiment with factor 1 random, 2 nested in 1 and 3 fixed. In this situation the paired observations comprising the bilinear form are not independent. The experiment was replicated 500 times for each simulation. The diagnostic has the form

$$\sum_{j} (\bar{Y}ijk. - \bar{Y}i.k.)(\bar{Y}ijk. - \bar{Y}i.k.)/(a_3 - 1).$$

Two cases were considered to determine how well the derived distribution fits correlated diagnostics from the diagnostic table. For case 1, $a_1 = 2$, $a_2 = 5$ and $a_3 = 3$. For this case there were three diagnostics per experiment for θ_{12}. Case 2 differed from case 1 in that a_3 was increased to 4. Both cases were generally similar. The maximum difference for the theoretical distribution in both cases ranged between 0.02 and 0.06, as shown in Table 2.1.

The difference between the theoretical and simulated numbers for the critical values of 0.01, 0.05, 0.10, 0.90, 0.95, and 0.99 ranged between 0.002 and 0.039, with the maximum difference occuring in the center of the distribution. The theoretical numbers were conservative, as in the independent case.

2.10 Tabulated Cumulative Distribution for the Diagnostics

Cumulative percentile points of the covariance distribution for ρ ranging between -0.7 to 0.9 in increments of 0.1, for sample size N of between 3 to 10, 15, 20, 25, 30, 40, 50, and for variance equal to one are contained in Grynovicki (1990). Due to the restriction of positive definitness, this range of parameters for ρ and N should be sufficient for most designs. Specifically, this table gives the value of C_{crit} such that $p(C \leq C_{crit}) = \alpha$ for $\alpha = 0.01, 0.05, 0.10, 0.90, 0.95$, and 0.99. C is a bilinear form from a bivariate normal with correlated paired observations.

TABLE 2.1
Calculated Kolmogorov-Smirnov One Sample Statistic, D, and Probability Differences at Critical Values for Simulation and Theoretical Distribution of Variance Component Diagnostics for Various Values of ρ from Non-Independent Sample Pairs

$a_2 = 3$

		Difference at Critical Probabilities α					
ρ	D	.01	.05	.10	.90	.95	.99
-.2	0.045	0.004	0.028	0.036	0.037	0.029	0.007
-.1	0.047	0.005	0.021	0.025	0.027	0.022	0.004
-.0	0.035	0.005	0.012	0.032	0.032	0.017	0.003
-.1	0.028	0.007	0.023	0.024	0.020	0.012	0.002
-.2	0.034	0.004	0.016	0.024	0.025	0.020	0.003
-.3	0.041	0.003	0.031	0.035	0.036	0.027	0.006
-.4	0.037	0.006	0.013	0.033	0.034	0.019	0.005
-.5	0.030	0.007	0.021	0.019	0.019	0.013	0.003
-.6	0.027	0.006	0.024	0.021	0.020	0.015	0.004
-.7	0.050	0.002	0.029	0.027	0.028	0.021	0.003
-.8	0.037	0.008	0.013	0.017	0.015	0.009	0.001
				$a_2 = 4$			
-.2	0.043	0.004	0.029	0.041	0.036	0.031	0.009
-.1	0.045	0.003	0.027	0.036	0.030	0.034	0.007
-.0	0.041	0.003	0.033	0.039	0.029	0.015	0.006
-.1	0.058	0.008	0.028	0.038	0.042	0.028	0.005
-.3	0.038	0.005	0.017	0.026	0.027	0.021	0.002
-.4	0.047	0.002	0.020	0.023	0.033	0.024	0.003
-.5	0.034	0.005	0.019	0.021	0.025	0.019	0.004
-.6	0.028	0.009	0.023	0.021	0.019	0.012	0.003
-.7	0.035	0.006	0.008	0.015	0.020	0.011	0.005
-.8	0.048	0.006	0.015	0.021	0.013	0.008	0.001

2.11 Illustrated Example Using Eye Glass Manufacturing Experiment

As an illustration of the diagnostic technique in comparison with its cumulative distribution, the diagnostic from an experiment previously examined by Green (1987) concerning eye glass manufacturing will be examined. The data for this experiment are presented in Table 2.2. Factor 1 (run) is random at five levels, factor 2 (pot) is random at two levels, and is nested in run, factor 3 (journey) is fixed at five levels, and factor 4 (period) is fixed at three levels. Factors 1, 3, and 4 are crossed.

In the previous analysis, Green clearly determined that runs 2 and 5 were highly variable and that pot 2, in journeys 2, 4, and 5 was clearly different from the rest of the data. The journey 2, between pot difference is extreme, and the journey 4 and 5, pot 2 values were from a different type of glass than all other responses.

Two diagnostic tables will be re-evaluated and are given in Tables 2.3 and 2.4. Table 2.3 represents the covariance $\sum_{j}(\bar{Y}ij.t. - \bar{Y}i..t.)(\bar{Y}ij.t.^{*} - \bar{Y}i..t.^{*})/(a_2 - 1)$ or, in Green's notation, $C(i,2/tt^{*})$. The variance covariance structure of $(\bar{Y}i1.t., \bar{Y}i2.t., \bar{Y}i1.t.^{*}, \bar{Y}i2.t^{*}.)$ is

$$V = \begin{pmatrix} aI + bJ_2J_2' & cI + dJ_2J_2' \\ cI + dJ_2J_2' & aI + bJ_2J_2' \end{pmatrix}, \text{ in which}$$

$$a = \phi_{12} + \phi_{123}/5 + \phi_{124} + \phi_{1234}/5,$$

$$b = \phi_{1} + \phi_{13}/5 + \phi_{14} + \phi_{134}/5,$$

TABLE 2.2

Glass Manufacture Data

		Pot					
		1			2		
		Period			Period		
		1	2	3	1	2	3
		47	56	100	52	61	88
		55	89	93	49	62	97
	1	35	57	56	34	60	72
		78	67	113	47	93	118
		33	40	128	16	29	130
		52	66	36	65	80	40
		21	61	49	122	97	79
	2	31	39	25	45	54	72
		43	72	52	109	120	80
		37	51	67	67	85	63
		50	61	60	75	139	130
		33	27	49	46	58	63
Run	3	24	39	24	15	33	39
		18	18	43	22	16	19
		28	42	28	27	19	22
		24	34	43	46	66	24
		24	49	42	40	117	105
	4	21	21	51	30	28	34
		21	69	48	36	64	53
		76	48	42	39	60	78
		31	54	40	19	93	36
		34	24	46	16	12	2
	5	120	122	120	33	58	107
		109	119	120	25	63	90
		69	49	60	34	43	30

TABLE 2.3

Diagnostics $\sum_j (\bar{Y}_{ij.t.} - \bar{Y}_{i..t.})(\bar{Y}_{ij.t.^*} - \bar{Y}_{i..t.^*})/(a_2 - 1)$

		t^* = 1	t^* = 2	t^* = 3	
t	1	50.00	4.00	-15.00	i = 1
	2		0.32	-1.20	
	3			4.50	
		t^* = 1	t^* = 2	t^* = 3	
t	1	1003.50	658.60	470.40	i = 2
	2		432.20	308.70	
	3			220.50	
		t^* = 1	t^* = 2	t^* = 3	
t	1	20.50	49.90	44.20	i = 3
	2		121.70	107.60	
	3			95.20	
		t^* = 1	t^* = 2	t^* = 3	
t	1	12.50	57.00	34.00	i = 4
	2		259.90	155.00	
	3			92.50	
		t^* = 1	t^* = 2	t^* = 3	
t	1	1113.90	467.30	571.10	i = 5
	2		196.00	239.60	
	3			292.80	

TABLE 2.4

Diagnostics $\sum_{T}(\bar{Y}_{1kt.} - \bar{Y}_{..kt.})(\bar{Y}_{1k^*t.} - \bar{Y}_{..k^*t.})/(a_1 - 1)$

	$k^*=$	1	2	3	4	5	
	1	250.7	207.1	-247.2	-61.8	-136.5	
	2		336.4	-131.6	221.3	-41.6	
k	3			497.5	357.8	124.8	t = 1
	4				620.1	75.5	
	5					236.3	

	$k^*=$	1	2	3	4	5	
	1	362.1	-312.8	29.1	-371.0	-107.4	
	2		799.9	-488.5	153.6	174.1	
k	3			631.2	416.2	-41.3	t = 2
	4				1009.6	312.6	
	5					229.7	

	$k^*=$	1	2	3	4	5	
	1	1012.7	350.0	-330.4	5.5	337.1	
	2		676.0	-488.6	41.1	763.8	
k	3			1032.3	888.3	103.4	t = 3
	4				1286.9	968.3	
	5					1530.2	

$$c = \phi_{12} + \phi_{123}/5, \text{ and}$$

$$d = \phi_1 + \phi_{13}/5.$$

Transforming the $\bar{Y}$s with the Helmert matrix would result in the transformed data having a variance covariance matrix of V with b and d set to zero. The variance of the transformed data would be $\phi_{12} + \phi_{123}/5 + \phi_{1234}/5 = 260$, the covariance $\phi_{12} + \phi_{123}/5 = 210$, and ρ is 0.8, based on the variance covariance estimates given in Green (1987). For these diagnostics, since N is 2, double the variances and use the 95% critical value with N = 2 and $\rho = 0.8$. The 95% confidence interval, [-98, 1705.6], is narrower than the 3σ criteria used previously. Due to the large variance of cell means for this table, no outliers were identified. This is consistent with the previous results. The high variability of runs 2 and 5, and low variability of run 1 is noticeable.

For Table 2.2, the variance of the cell means is $\phi_{12} + \phi_{123}/2 + \phi_{124}/2 + \phi_{1234}/2 + \phi_1 + \phi_{13} + \phi_{14} + \phi_{134}$. Table 2.4 represents the covariance $C(t,1/kk^*) =$

$$\sum_I (\bar{Y}i.kt. - \bar{Y}..kt.)(\bar{Y}i.k^*t. - \bar{Y}..k^*t.)/(a_1 - 1).$$

The variance-covariance structure of the cell means comprising this bilinear form is

$$V = \begin{pmatrix} aI & cI \\ cI & aI \end{pmatrix},$$

in which $a = \phi_1 + \phi_{12}/2 + \phi_{13} + \phi_{123}/2 + \phi_{14} + \phi_{124}/2 + \phi_{134} + \phi_{1234}/2$ and

$$c = \phi_1 + \phi_{12}/2 + \phi_{14} + \phi_{124}/2.$$

Using the estimates of the variance components found in Hocking (1989), a = 701.17, and b = 92.43. Thus, the estimated correlation of the independent paired cells of different journey conditions for a given period is 0.132. Using the distribution theory, one can obtain an estimate of the 95% confidence interval [-442, 716]. Based on this interval, one can see that period 2 journey 2 and 3 covariance is small and period 3 journey (3, 4), (2, 5) and (4, 5) covariances were outside the 95% confidence interval specified above. The low covariance in period 2 may be due to run 5, pot 2, period 3, journey 2, which was identified by Green (1987) as an outlier. The large covariances are because of run 1, journey 5 and run 5, journey 2, period 2 and run 5, journey 3, 4. It should be noted that in run 5, all responses were from different furnaces than were used in the other runs.

2.12 Conclusions

The distribution of the diagnostics for a bilinear form when the sample pairs are independent and not independent has been developed, tabulated, and validated. This theory has been extended to the diagnostic tables for all random and mixed designs. For the special case when N = 2, it has been shown that the bilinear form for non-independent sample pairs is equivalent to the independent case with the variance doubled.

BIBLIOGRAPHY

Abramowitz, M., & Stegun, I. A. (Eds.). (1964, December). Handbook of mathematical functions with formulas, graphs, and mathematical tables (Applied Mathematical Series No. 55). Washington, DC: U.S. Government Printing Office.

Green, J. W. (1987). Diagnostic procedure for variance components for both large and small designs. Manuscript submitted for publication.

Grynovicki, J. O. (1989). Variance component estimation diagnostics and associated distribution theory for all random and mixed designs application to repeated measures. Doctoral dissertation, University of Delaware.

Grynovicki, J. O., & Green, J. W. (1990). Diagnostics and distribution theory for all random and mixed experimental designs. (TM- -) Aberdeen Proving Ground, MD: U.S. Army Laboratory Command, Human Engineering Laboratory.

Hirschfeld, H. O. (1937). The distribution of the ratio of covariances estimates in samples drawn from normal bivariate populations. Biometrika, 29, 65-79.

Hocking, R. R. (1985). The analysis of linear models, Monterey, CA: Brooks-Cole Publishing Company.

Hocking, R. R., Green, J. W., & Bremer, R. H. (1989). Estimation of fixed effects and variance components in mixed factorial models including model-based diagnostics. Technometrics, 31, 227-240.

Mahalanobis, P. C., Bose, R. C., & Roy, S. N. (1937). Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. Sankhya, 3, 1-40.

Morrison, J. M. (1987). [Class notes from Real Analysis]. University of Delaware.

Pearson, K., Jeffery, G. B., & Elderton, E. M. (1929). On the distribution of the first product moment-coefficient in samples drawn from an indefinitely large normal population. Biometrika, 21, 164-201.

Press, S. J. (1967). On the sample covariance from a bivariate normal distribution. Annals of the Institute of Statistical Mathematics (Tokyo), 19, 355-361.

Robbins, H., & Pitman, E. J. G. (1949). Application of the Method of Mixtures to quadratic forms in normal variates. Annals of Mathematical Statistics, 20, 552-560.

Searle, S. R. (1971b). Linear models. New York: John Wiley & Sons.

Slater, L. J. (1965, December). Confluent hypergeometric functions. In M. Abramowitz, & I. A. Stegun (Eds.), Handbook of mathematical functions with formulas, graphs, and mathematical tables (Applied Mathematical Series No. 55). Washington, DC: U.S. Government Printing Office.

Wheeden, R. L., & Zygmund, A. (1977). Measure and integral: An introduction to real analysis. New York: Marcel Dekker.

Wishart, J., & Bartlett, M. S. (1932). The distribution of second order moment statistics in a normal system, Proceedings of the Cambridge Philosophical Society, 28, 455-459.

Numerical Estimation and Properties of the Source Density Function

Charles E. Hall, Jr.
Research Directorate
Research, Engineering, and Development Center
U.S. Army Missile Command
Redstone Arsenal, AL 35898-5248

ABSTRACT

The Source Density Function is a four-parameter class of one-sided probability density functions. In order to exploit the Source Density Function's flexibility in shape, programs were developed to estimate the parameters which maximize the log-likelihood function for a given data set.

INTRODUCTION

A brief review of the Source Density Function (SDF) is presented here a rigorous development was done by Lehnigk[1]. The SDF, f(x,P), is generated from a delta function initial condition solution of the generalized Feller equation.

$$f(x,P) = \beta\, b^{-\beta}\, x^{-(p-\beta+1)/2}\, z^{(p+\beta-1)/2}\, I_q[2(xz/b)^{\beta/2}]\, \exp[-b^{-\beta}(x^{\beta}+z^{\beta})] \qquad (1)$$

$$P = (\, z \;\; b \;\; p \;\; \beta\,)'$$

$$z>0; \quad b>0; \quad p<1; \quad \beta>0$$

$I_q(\cdot)$ is the modified Bessel function of the first kind, where $q = -1 + (1-p)/\beta > -1$. The vector P is composed of the four parameters which are calculated so that the log-likelihood function is maximized. A data set of observations is formed, which is composed of ordered pairs of the observation variable x_ν, and the relative frequency of that observation f_ν. The data set, $\{(x_\nu,f_\nu)|\nu=1,2,\ldots,n$ with f_0 and $f_n \neq 0\}$, is used with f(x,P) to form the log-likelihood function $\phi(P)$.

$$\phi(P) = \sum_{\nu=1}^{n} f_\nu \ln(\, f(x_\nu,P)\,) \qquad (2)$$

It should be noted that as $z \to 0$ both equations (1) and (2) approach the Hyper-Gamma density and log-likelihood functions for $\lambda=0$ [2]. This will be refered to as the Hyper-Gamma limit of the SDF.

A transformation of the parameters is useful in simplifying the equations.

$$\sigma = z^{\beta/2} \tag{3}$$

For a maximum of the log-likelihood function, the requirement exists that all of the derivatives of $\phi(P)$ must equal zero. These equations place a further restriction on σ, and it allows the elimination of the parameter b from the equations.

$$b^{\beta} = (B(\beta) - \sigma^2) / (1+q) \tag{4}$$

$$B(\beta) = \sum_{\nu=1}^{n} f_\nu \exp(\beta \rho_\nu) \tag{5}$$

$$\rho_\nu = \ln(x_\nu) \tag{6}$$

For $b>0$, it is required that $B(\beta) - \sigma^2 > 0$, thus $0 < \sigma < \sqrt{B(\beta)}$. Equation (4) allows the elimination of b from (2), thus $\phi(P)$ is a three-parameter equation.

$$\phi(\sigma,\beta,\mu) = \ln \beta + \mu \ln(\mu/(B(\beta)-\sigma^2)) + (\mu\beta - 1)C - \mu \frac{B(\beta)+\sigma^2}{B(\beta)-\sigma^2} + \sum_{\nu=1}^{n} f_\nu \ln(S_{\mu-1}(r_\nu)) \tag{7}$$

$$C = \sum_{\nu=1}^{n} f_\nu \rho_\nu \tag{8}$$

$$S_{\mu-1}(r) = (2/r)^{\mu-1} I_{\mu-1}(r) = \sum_{k=0}^{\infty} (r/2)^{2k} / k!\, \Gamma(k+\mu) \tag{9}$$

$$r_\nu = 2\mu\sigma (B(\beta)-\sigma^2)^{-1} \exp(\beta\rho_\nu/2) \tag{10}$$

$$\mu = 1 + q \tag{11}$$

Equations (4-11) form the starting point for the numerical estimation of the source density function parameters σ, β, and μ.

METHODS FOR PARAMETER ESTIMATION OF THE SOURCE DENSITY FUNCTION

Initial attemps at parameter estimation of the SDF were based on the simultaneous solution of the derivative equations of (7) set equal to zero. These equations had the following form.

$$0 = -\sigma + \sum_{\nu=1}^{n} f_\nu \, (S_q(r_\nu))^{-1} \, \frac{d\, S_q(r_\nu)}{dr_\nu} \exp(\beta\rho_\nu/2) \tag{12}$$

$$0 = (B(\beta)-\sigma^2)(1+\mu\beta C) - \mu\beta \sum_{\nu=1}^{n} f_\nu\rho_\nu\exp(\beta\rho_\nu) - \sigma\sum_{\nu=1}^{n} \frac{f_\nu\rho_\nu}{S_q(r_\nu)} \frac{dS_q(r_\nu)}{dr_\nu} \exp(\beta\rho_\nu) \tag{13}$$

$$0 = \beta C + \ln(\mu \, (B(\beta) - \sigma^2)^{-1}) + \sum_{\nu=1}^{n} \frac{f_\nu}{S_q(r_\nu)} \frac{dS_q(r_\nu)}{dq} \tag{14}$$

A three dimensional application of the Newton-Raphson method was used. The functions on the right side of the equal sign of equations (12-14) were used to form a vector, $F(\sigma,\beta,\mu)$ and a 3x3 derivative matrix of $F(\cdot)$ was numerically calculated. This matrix was inverted and premultiplied the negative of $F(\cdot)$ to yield a change vector for the three parameters. This method failed to produce useful results due to the complexity of the $\phi(\cdot)$ function which typically had differences between the σ-derivative and β-derivative functions that typically spanned 10 or more orders of magnitude. The derivative based approach was abandoned in favor of direct optimization methods.

Direct optimization of various log-likelihood functions by Powell's Method have been successful [2,3,4,5], so this technique was applied to equation (7). Initial runs, with the starting point close to the actual parameters that were used to generate the data sets, were successful. But as the starting point was moved further away from the solution, Powell's algorithm ran into difficulties due to its inability to deal with the β-σ boundary generated by $B(\beta)-\sigma^2>0$, (and the flatness of $\phi(\cdot)$).

Powell's method is an unconstrained minimization algorithm. To change a maximum into a minimum, the function is multiplied by -1. In this paper all equations will be presented as they were derived, and it is understood that the log-likelihood function is multiplied by -1 in the computer programs. The next alteration required is to change Powell's algorithm into a constrained minimization. For the Log-Normal, generalized Gumbel, and the Hyper-Gamma distributions all of the constraints were implemented in the calculation of the log-likelihood function. If in the function subroutine, it was detected that a parameter had gone outside the allowable region, then the function would force the offending parameter into the allowed domain. This proved satisfactory

since for these distributions, all of the constraints and the direction vectors for Powell's algorithm were parallel to the coordinate axis system, but for the SDF this was not the case on the β-σ boundary. A modification to the Powell algorithm was made in the minimum bracketing subroutine, MNBRK. If the function detected a parameter which was not in the allowable region a flag was set, this flag was a signal to NMBRK that a constraint had been crossed. MNBRK would then bisect the interval between the last good point and the desired point which had crossed the boundry, and then try this new point. This procedure is repeated until the test point was in the allowable region. This improved the region of convergence, but it still remained too limited.

To further modify Powell's algorithm to get a better convergence criterion, it was necessary to examine the structure of the log-likelihood function for the source density function. Figures 1 and 2 show cuts of the log-likelihood function as it varies with σ (β and μ fixed) and μ (σ and β fixed) with the two fixed parameters set at the solution values. The scales on these plots are to demonstrate the flatness of the function. These indicate that the $\phi(\cdot)$ function is a well-behaved parabolic type function, and this continues even when the fixed parameters are set at non-solution values, (of course with its extremum value decreased). Unfortunately, this is not the case when $\phi(\cdot)$ is made a function of β, with μ and σ set at the solution values, (shown in Figure 3). During the investigation it was seen that the left-hand peak of Figure 3 was the extremum, while the right-hand was a false extremum. If the μ or σ parameter varied off of the solution value, the two peaks moved towards each other and the left-hand peak was absorbed into the right-hand peak. This demonstrates the existence of a ridge that connects the two maximums of Figure 3 together. This ridge must be followed by Powell's algorithm to locate the maximum. Figure 4 shows a typical ridge in σ-β-μ space.

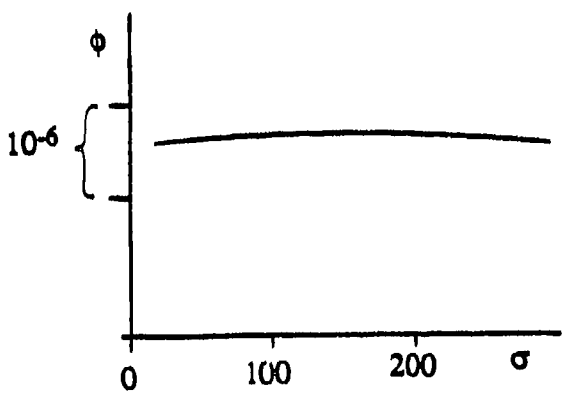


Figure 1. $\phi(\sigma)$, β and μ constant.

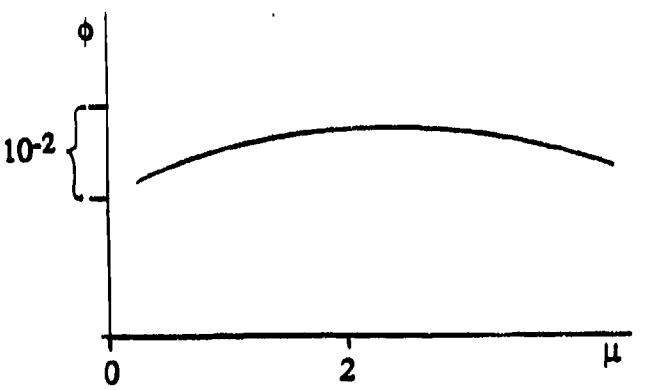


Figure 2. $\phi(\mu)$, σ and β constant.

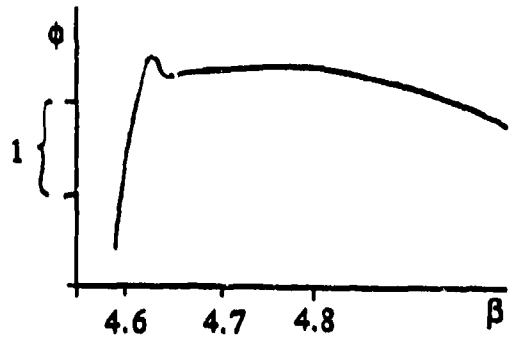


Figure 3. $\phi(\beta)$, σ and μ constant.

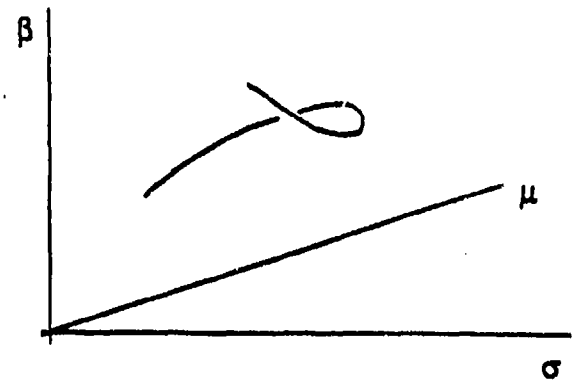


Figure 4. Powell's trajectory in σ-β-μ space.

At first inspection it appears that this ridge was exactly what Powell was developed for, but there are problems with traveling along this ridge. The first difficulty is that the relative change in traveling along this ridge is approximately 1 part in 10^5 to 10^6, and it takes numerous iterations following the ridge. When the relative change along the ridge is divided by the number of iterations required for that journey, this average relative change is usually less than the termination criteria for Powell's method. Thus, Powell terminates the optimization on a false maximum. Two means were employed to alleviate this problem. First, the entire Powell subroutine package was rewritten to perform all calculations in double precision. The evaluation of the log-likelihood function was always performed in double precision to improve accuracy. With Powell's subroutines being in double precision, the termination criteria was improved, which helped to increase the range of convergence. To further increase the convergence area an amplifier function, equation (15), was applied to the log-likelihood function for a second pass after the termination criteria was satisfied on the first pass by the Powell subroutine package.

$$\psi = e^{10} \left(e^{10(\phi-\phi^*)} - 1 \right) \qquad (15)$$

ϕ^* was the final value of ϕ from the first pass of Powell's algorithm. The second pass of Powell was used to maximize the ψ function. The amplifier function increases the slope of the function, while eliminating the large dc-offset. From earlier work with this amplifier, it was observed that the termination criteria was effectively changed from 1 part in 6×10^6 (Powell in single precision, ϕ and ψ calculated in double precision) to 1 part in 10^{10}. The actual amount of increase in the effective termination criteria on ϕ is dependent on the difference in ϕ and ϕ^*, a small difference yielded a better termination criteria (1 part in 10^{11}) while a large difference lessened the termination criteria (1 part in 10^9). Unfortunately, these modifications did not fully solve the problem, but they

did help. Occasionally, the function was so flat on the ridge that even with the amplifier function, Powell's termination criteria was satisfied. It appears that increasing the gain of the amplifier would be of assistance, but Powell's trajectory could be close to a boundary thus causing a large change which would result in an overflow. Restarting the amplifier with a new ϕ^* did help to extend the range of convergence; thus Powell's algorithm was running with three passes, one plain and two with the amplifier.

Even with this, the convergence range did not equal the allowable space. In some regions, the Powell algorithm would "lock" on to a false maximum. At some of these false maximums, a plot of ϕ as a function of one parameter σ, β, or μ would show a maximum, but a ridge did led away from this point in a direction oblique to the coordinate axis. During initialization, the Powell subroutine was given a set of direction vectors, which spanned the space, and Powell's method searched for successive extremum along these direction vectors. The direction vectors were changed, allowing an escape from the original false maximum but it would usually fall prey to another. Similarly, Powell's algorithm at times needed to track along a curved ridge or boundary, but this would trigger a similar false maximum. To get past the false maximum problem, a steepest descent subroutine package was written. This method was successful in finding the ridge, but it failed once on the ridge, due to the flatness.

A variable transformation was then tried. Changing to λ did again help extend Powell's range,

$$e^{\lambda} = B(\beta) - \sigma^2 \qquad (16)$$

but this did not fully solve the problems.

Figure 5 is the computer output from four runs. The λ,β and μ values are the initial values. The λ, β, μ, b and σ, z are the final Powell estimates. All four runs did converge.

CONCLUSIONS

Application of Powell's method in three passes does produce accurate estimates of the parameters of the Source Density Function. The major drawback is the requirement of a starting point that lies in the convergence zone of the global maximum. In previous programs which utilized the maximum log-likelihood principle with distribution such as, Log-Normal, generalized Gumbel, and Hyper-Gamma, the moment estimates became the starting point for Powell's method. The moment estimates for the Source Density Function require simultaneous solution of four

```
 SOURCE DENSITY FUNCTION CALCULATIONS
Data file is  \SDFDATA\SYNEX1S.DAT
Lambda =    .800000E+001
Beta   =    .300000E+001
Mu     =    .300000E+001
*****************************************************
THE POWELL ESTIMATES FOR THE SOURCE DENSITY FUNCTION
Lambda =     .120232E+002
Beta   =     .462324E+001
Mu     =     .207676E+001
b      =     .115021E+002

Sigma  =    .343231E+003
z      =    .124994E+002
-----------------------------------------------------

Lambda =    .800000E+001
Beta   =    .300000E+001
Mu     =    .300000E+001
*****************************************************
THE POWELL ESTIMATES FOR THE SOURCE DENSITY FUNCTION
Lambda =     .120232E+002
Beta   =     .462324E+001
Mu     =     .207676E+001
b      =     .115021E+002

Sigma  =    .343231E+003
z      =    .124994E+002
-----------------------------------------------------

Lambda =    .000000E+000
Beta   =    .300000E+001
Mu     =    .300000E+001
*****************************************************
THE POWELL ESTIMATES FOR THE SOURCE DENSITY FUNCTION
Lambda =     .120229E+002
Beta   =     .462313E+001
Mu     =     .207679E+001
b      =     .115019E+002

Sigma  =    .343193E+003
z      =    .124995E+002
-----------------------------------------------------

Lambda =    .800000E+001
Beta   =    .900000E+001
Mu     =    .300000E+001
*****************************************************
THE POWELL ESTIMATES FOR THE SOURCE DENSITY FUNCTION
Lambda =     .120241E+002
Beta   =     .462381E+001
Mu     =     .207668E+001
b      =     .115027E+002

Sigma  =    .343338E+003
z      =    .124992E+002
```

Figure 5. Four sample runs.

nonlinear equations, and this has proved to be more difficult than the maximum log-likelihood estimate.

REFERENCES

[1] "A Source Density Function"; Lehnigk, S. H.; "*Math. Meth, in the Appl. Sci.*"...

[2] "Parameter Estimation Algorithms for the Hyper-Gamma Distribution Class"; Dudel, H. P., Hall, C. E. Jr., and Lehnigk, S. L.; *Communications in Statistics*, 18(3), 1135-1153, 1989

[3] "An Extreme Value Type I Distribution, Moment and maximum Likelihood Estimation"; Hall, C. E. Jr., and Lehnigk, S. H.; *Journal of the Royal Statistical Society*; to be published.

[4] "Log-Normal Distribution Maximum Likelihood Parameter Estimation Algorithms"; Dudel, H. P., Hall, C. E. Jr., and Lehnigk, S. H.; Tech Report RD-RE-89-1, U.S. Army Missile Command, ATTN:AMSMI-RD-RE-OP, Redstone Arsenal, AL 35898-5248

[5] "*Numerical Recipes': The Art of Scientific Computing*"; Press,W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T.; Cambridge University Press, c1985.

THE HUNTER PROBLEM IN A RANDOM FIELD OF OBSCURING ELEMENTS

Shelemyahu Zacks
State University of New York at Binghamton
and
M. Yadin
Technion-Israel Institute of Technology

ABSTRACT

A hunter attempts to detect and kill targets within a field of obscuring elements, which are randomly dispersed (trees in a forest). The targets move along paths in the field, which are partially obscured by the random elements. When a target enters a visible segment of a path it takes t_0 [seconds] to detect it, and t_1 [seconds] to attempt destroying it. If such a trial is not successful, other independent trials can be performed as long as the target is visible. The number of shooting trials that can be attempted depends on the number and lengths of the visible portions of the path. Lower and upper bounds for the probability of destroying a target are determined by using the methods of random visibility measures previously developed by the authors.

Key Words: *Poisson Shadowing process, Bernoulli Trials, Visibility Probabilities, τ-reduced measure of Visibility, Detection Probability, Hitting Probability.*

*Partially supported by Contract DAAG29-83-K-0176 and Contract DAAL03-89-K-0129 with the U.S. Army Research Office.

0. Introduction

A hunter is trying to detect and hit a target in a forest. Suppose that a target is moving along a path in the forest and the hunter is located among the trees at some distance from the path. The path is only partially visible to the hunter; the invisible (shadowed) portion of the path is obscured by the trees which are dispersed randomly between the hunter and the path. A target can be detected by the hunter if at least a certain part of it is visible. After detection of a target, the hunter starts shooting. The target continues to move along the path in the same pace. During each shooting trial the target crosses a length of τ of the path. Thus the number of shooting trials in each visible segment depends on the length of the segment. The shooting trials stop either when the target is hit or when it enters an invisible portion of the path. When the target enters another visible segment, it has to be detected again. For simplicity we assume that the shooting trials are Bernoulli, with probability of failure q, $0 < q < 1$.

The problem of target hunting can be treated as a two or three dimensional shadowing problem. Two dimensional random shadowing problems were previously studied by Chernoff and Daly [1]. Likhterov and Gurin [2], Yadin and Zacks [3,4]. The methodology developed in the present paper is also applicable to three dimensional versions of the above problem. For example, if a hunter tries to shoot down a helicopter whose flying course is partially obscured by crowns of trees. The three dimensional shadowing problem was previously studied by Yadin and Zacks [5].

In the present study we develop approximations for (a) the probability of detection; (b) the probability distribution of the maximal number of shooting trials N; and (c) the probability of survival of the target. We also provide numerical examples to illustrate the goodness of these approximations.

1. The Model, Measures of Visibility and Failure Probabilities

Suppose that the hunter is located at the origin, $\underset{\sim}{0}$, and let C denote the path of the target. C is assumed to be a smooth star shaped curve, defined by a piece-wise differentiable function $r(s), s_L \leq s \leq s_U$, representing the distance from $\underset{\sim}{0}$ to C in orientation s. The polar coordinates of a point $\underset{\sim}{P}_s$ on C are $(r(s),)$. The end-points of C are $\underset{\sim}{P}_{s_L}$ and $\underset{\sim}{P}_{s_U}$. The length of C is

$$L = \int_{s_L}^{s_U} l(s)ds \tag{1.1}$$

where

$$l(s) = [r^2(s) + (\frac{d}{ds}r(s))^2]^{1/2}$$

The trees in the forest are presented by random disks dispersed in a region between $\underset{\sim}{0}$ and C. Each random disk is characterized by coordinates (ρ, θ, y), where (ρ, θ) are the polar coordinates of its center and y is its diameter. The coordinates (ρ, θ, y) belong to a set S in R^3 satisfying conditions which assure that $\underset{\sim}{0}$ is not covered and C is not intersected by random disks. Let B be the Borel σ-field on the sample space S, and let $N\{B\}$ designate the number of disks whose coordinates belong to a set B of B. We assume that, for each $B \varepsilon B$, $N\{B\}$ is a random variable having a Poisson distribution with mean

$$\nu\{B\} = \lambda \int_B \int \int H(d\rho, d\theta) dG(y|\rho, \theta). \tag{1.2}$$

where $G(y|\rho,\theta)$ is the conditional CDF of y, given (ρ,θ), and $H(d\rho,d\theta)$ is a σ-finite measure of (ρ,θ). Such a random field of disks is called a Poisson random field.

A point $\underset{\sim}{P}_s$ on C is said to be visible if the line segment $\overline{\underset{\sim}{O}\underset{\sim}{P}}_s$ is not intersected by any random disk. A point which is not visible is in a shadow. The measure of total visibility on C is defined as

$$V = \int_{s_L}^{s_U} I(s)l(s)ds \quad , \tag{1.3}$$

where $I(s) = 1$ if $\underset{\sim}{P}_s$ is visible, and $I(s) = 0$ otherwise. Notice that V is a random variable representing the total length of the visible portion of C. V is a sum of a random number, M, of visible segments of C having random length $X_1, X_2, \ldots, X_M$; i.e.

$$V = \sum_{i=1}^{M} X_i \quad . \tag{1.4}$$

A target is detected only if there exists at least one visible segment of length greater than the minimal path length τ_o required for identifying the target. In order to develop a formula for the probability of detecting a target, we introduce the notion of τ-reduced visibility measure, $V(\tau)$, which is the total length of visible segments, each one reduced by τ units, i.e.,

$$V(\tau) = \sum_{i=1}^{M} (X_i - \tau)_+ \tag{1.5}$$

where $a_+ = \max(a, 0)$. The probability that a target is not detected is

$$p_o(\tau_o) = Pr\{V(\tau_o) = 0\}. \tag{1.6}$$

On the other hand, the probability that C is completely visible is

$$p_1 = Pr\{V(\tau) = L - \tau\}, \text{ for all } 0 \le \tau \le L. \tag{1.7}$$

Indeed, when C is completely visible, $M = 1$ and $X_1 = L$. Let N denote the number of shooting trials, after detecting a target. If a single shooting trial requires a setment of length τ to be completely visible, then

$$N = \sum_{i=1}^{M} [(X_i - \tau_o)_+/\tau] \, , \tag{1.8}$$

where [a] is the maximal integer not exceeding a. Notice that

$$\frac{1}{\tau}\sum_{i=1}^{M} (X_i - \tau_o - \tau)_+ \le N \le \frac{1}{\tau}\sum_{i=1}^{M} (X_i - \tau)_+ \tag{1.9}$$

Hence, according to (1.5) and (1.9),

$$V(\tau_1)/\tau \le N \le V(\tau_o)/\tau \quad , \tag{1.10}$$

where $\tau_1 = \tau_0 + \tau$.
If the probability of failure in each shooting trial is q, and the shooting trials are independent (Bernoulli), the number of shooting trials required until the first success, J, is distributed geometrically. Accordingly, the probability of failure (not hitting the target) is $Q = E\{q^N\}$. Thus, according to (1.10), lower and upper bounds for Q are, respectively, Q_0 and Q_1, where

$$Q_i = E\{q^{V(\tau_i)/\tau}\} \quad , i = 0, 1 \tag{1.11}$$

Notice that Q_i is the value of the MGF of $V(\tau_i)$ at the point $t = (\log q)/\tau$.

2. The Moments and Moment Generating Function of $V(\tau)$.

For the sake of determining the moments of $V(\tau)$ we introduce the following definition of this measure,

$$V(\tau) = \int_{s_{L,\tau}}^{s_{U,\tau}} I_\tau(s) l(s) ds \quad , \tag{2.1}$$

where $I_\tau(s) = 1$ if a segment of C of length τ, centered at $(r(s), s)$ is completely visible, and $I_\tau(s) = 0$ otherwise. $s_{L,\tau}$ and $s_{U,\tau}$ are direction coordinates of points within C, of distance $\tau/2$ along C from s_L and s_U respectively. More formally, let

$$L(s) = \int_{s_L}^{s_U} l(y) dy. \tag{2.2}$$

Then, $s_{L,\tau} = L^{-1}(\tau/2)$ and $s_{U,\tau} = L^{-1}(L - \tau/2)$.
The n-th moments of $V(\tau)$ is thus

$$\begin{aligned} \eta_n(\tau) &= E\{(\int_{s_{L,\tau}}^{s_{U,\tau}} I_\tau(s) l(s) ds)^n\} \\ &= n! \int_{A_{n,\tau}} \cdots \int E\{\prod_{i=1}^{n} I_\tau(s_i)\} \prod_{i=1}^{n} l(s_i) ds_i. \end{aligned} \tag{2.3}$$

The set $A_{n,\tau}$ is the simplex

$$A_{n,\tau} = \{(s_1, \ldots, s_n);\ s_{L,\tau-} < s_1 < \ldots < s_{n-} < s_{U,\tau}\} \quad . \tag{2.4}$$

Furthermore, $E\{\prod_{i=1}^{n} I_\tau(s_i)\}$ is the probability that the union of n segments of C, each one of length τ, centered at n points having direction coordinates $s_1 < \ldots < s_n$, is completely visible. This probability is designated by $p_n(s_1, \ldots, s_n; \tau)$. Thus the n-th moment of $V(\tau)$ is

$$\eta_n(\tau) = n! \int_{A_{n,\tau}} \cdots \int p_n(s_1, \ldots, s_n; \tau) \prod_{i=1}^{n} l(s_i) ds_i. \tag{2.5}$$

The method for determining $p_n(s_1, \ldots, s_n; \tau)$ and $\mu_n(\tau)$ is based on a general methodology developed by Yadin and Zacks [3,4] for the special case of $\tau = 0$ the modifications required for $\tau > 0$, are given in a Technical Report [6].

3. An Approximation to the CDF of $V(\tau)$

The cumulative distribution function (CDF) of $V(\tau)$ is a mixture of a two-point distribution concentrated on $\{0, L-\tau\}$ and a distribution concentrated on the interval $(0, L-\tau)$. For the purpose of presenting the approximation discussed below, we consider a normalized measure of visibility $W(\tau) = V(\tau)/(L-\tau)$, which is concentrated on [0,1]. The CDF of $W(\tau)$ can be represented as

$$F_\tau(w) = \begin{cases} 0 & , \text{ if } w < 0 \\ p_0(\tau) + (1 - p_0(\tau) - p_1)F^{*_\tau}(w) & , 0 \le w < 1 \\ 1 & , 1 \le w \quad . \end{cases} \tag{3.1}$$

If, for example, $G(y|\rho, \theta)$ is absolutely continuous then $F^*_\tau(w)$ is an absolutely continuous CDF on $(0,1)$. Let $\mu_n(\tau)$ denote the n-th moment of $W(\tau)$. Obviously, $\eta_n(\tau) = (L-\tau)^n\mu_n(\tau), n = 1,2,\ldots$.
Furthermore, for $n = 1,2,\ldots$

$$\mu_n(\tau) = p_1 + (1 - p_0(\tau) - p_1)\int_0^1 w^n dF^*_\tau(w), \tag{3.2}$$

Applying the Dominated Convergence Theorem one immediately proves that $\lim_{n\longrightarrow\infty}\mu_n(\tau) = p_1$ for all $\tau \ge 0$.

Explicit expressions for $p_0(\tau)$ and $F^*_\tau(w)$ are not available. We apply here a beta approximation to $F^*_\tau(w)$ and provide a numerical approximation to $p_0(\tau)$. This type of mixed-beta approximation was applied also in [3,4,5]. As will be shown in Section 6, in some special cases, the first ten moments of $W(\tau)$ and of the mixed-beta approximation are very close. This indicates that in those cases one has a highly effective approximation. In cases where the moments are not in agreement better approximation should be attempted. The approximating beta-mixture CDF is given by the formula

$$\tilde{F}_\tau(w) = \begin{cases} 0 & , \text{ if } w < 0 \\ \tilde{p}_0(\tau) + (1 - \tilde{p}_0(\tau) - p_1)I_w(\alpha_\tau, \beta_\tau) & , 0 \le w < 1 \\ 1 & , \text{ if } 1 \le w \quad , \end{cases} \tag{3.3}$$

where $I_w(\alpha,\beta), 0 \le w \le 1, 0 < \alpha, \beta < \infty$, denotes the incomplete beta function ratio. The probability p_1 of complete visibility of the segment (S_L, S_u) of C is determined by the shadowing model, as shown later. The values of $\tilde{p}_0(\tau), \alpha_\tau$ and β_τ are determined by equating the formulae of the first three moments of $\tilde{F}_\tau(w)$ to those of $W(\tau)$, as shown in [3].

4. Bounds for the CDF of N and for Q

Inequality (1.10) yields lower and upper bounds for the CDF of N. Indeed, from (1.10),

$$F_{\tau_0}(\frac{\tau n}{L-\tau_0}) \le Pr\{N \le n\} \le F_{\tau_1}(\frac{\tau n}{L-\tau_1}) \tag{4.1}$$

The CDF's in (4.1) can be approximated by the mixed-beta CDF (3.3). According to (1.11), the lower and upper bounds, for the failure probability Q, are the value of the

MGF of $W(\tau_i) i = 0, 1$, at the point $t = \frac{1}{\tau}(L - \tau_i) \log q$. Let $G_\tau(t)$ indicate the MGF of $W(\tau)$. This function can be expressed in terms of the moments of $W(\tau)$ as

$$G_\tau(t) = 1 + p_1(e^t - 1) + \sum_{n=1}^{\infty} \frac{\mu_n^*(\tau)}{n!} t^n, \quad -\infty < t < \infty. \tag{4.2}$$

Since $\mu_n^*(\tau) \downarrow 0$ as n grows the infinite series in (4.2) converges faster than e^t, and therefore a small number of terms will often provide a good approximation. Another method of approximating $G_\tau(t)$ is by employing the MGF of the mixed-beta distribution (3.3) with $\tilde{p}(\tau), \alpha_\tau$ and β_τ.

5. Numerical Example

In the present section we provide an example which demonstrates numerically the results of the present paper. We consider the case of an arc C and annular strip S, which was discussed in Section 6.1. The parameters of this case are:

$$\theta_L = -\pi/2, s_L = -\pi/3, s_U = \pi/3, \theta_u = \pi/2, r = 1, w = .6, u = .4, \lambda = 5.$$

In addition, the diameters are uniformly distributed over the interval (.1, .5).

In Table 5.1 we present the first 10 moments of $W(\tau)$, for $\tau = 0(.1).4$. The corresponding moments of the mixed-beta distribution (3.3) are also given for comparison.

As shown in Table 5.1, the first ten moments obtained from the mixed-beta CDF, $\tilde{F}_\tau(w)$, differ from those of the correct distribution only at the 4th decimal place. This reveals an excellent approximation to the CDF of $W(t)$ by $\tilde{F}_\tau(w)$, in the case under consideration. In Table 5.2 we provide the parameters of the mixed-beta distributions associated with Table 5.1.

The values of $\tilde{p}_0(\tau)$ in Table 5.2, provide the mixed-beta approximations to the probabilities $p_0(\tau_0)$ of not detecting a target. This is obviously an increasing function of τ_0. Thus, in the present example, if $\tau_0 = .1, \tilde{p}_0(\tau_0) = .012$ while if $\tau_0 = .4, \tilde{p}_0(\tau_0) = .043$. $p_1 = .27$ is the probability of complete visibility along the path. Since the moments of the mixed-beta distributions $\tilde{F}_\tau(w)$ fitted so well those of $W(\tau)$, we replace $F_{\tau_i}(\frac{\tau n}{L-\tau_i})$ with $\tilde{F}_{\tau_i}(\frac{\tau n}{L-\tau_i}), i = 0, 1$. In Table 5.3 we present $\tilde{F}_{\tau_i}(\frac{\tau n}{L-\tau_i})$ for $\tau_i = 0(.1).4, \tau = .1$.

The values of $Q_i = E \exp\{t_i W(\tau_i)\}\}$ where $t_i = \frac{s_n - s_L - \tau_i}{\tau} \log(q)$ with $q = .8$, are also given in Table 5.3.

As seen in Table 5.3, if $\tau = .1$ and $\tau_0 = .1$ the lower bound of Q is .0967 and the upper bound for Q is .1273. If however, $\tau_0 = 0$ then $.0704 \leq Q \leq .0967$.

The bounds for the CDF of N are read from Table 5.3 in a similar manner. For example, if $\tau_0 = 0, \tau_1 = .1 + \tau_0 = .1$ then for $n = 6, .0435 \leq P\{N \leq 6\} \leq .0785$. If, $\tau_0 = .1$ then $\tau_1 = .1 + \tau_0 = .2$ and $.0785 < P\{N \leq 6\} \leq .1253$. Thus, from the first two columns of Table 5.3 we obtain that, when $\tau_0 = 0$, the expected number of trials, $E\{N\}$, is between 13.7 and 15.1.

τ \ n	1	2	3	4	5	6	7	8	9	10
0.0	.738 .738	.600 .600	.517 .517	.462 .463	.425 .425	.398 .398	.378 .378	.363 .363	.351 .351	.342 .342
0.1	.704 .704	.561 .561	.479 .479	.427 .427	.393 .393	.369 .369	.351 .351	.338 .338	.329 .329	.321 .321
0.2	.671 .671	.526 .526	.447 .447	.399 .399	.368 .369	.347 .347	.332 .332	.321 .321	.313 .313	.307 .307
0.3	.641 .641	.497 .497	.421 .421	.377 .377	.349 .349	.330 .331	.318 .318	.308 .309	.302 .302	.297 .297
0.4	.614 .614	.471 .471	.399 .399	.359 .359	.334 .334	.318 .318	.307 .307	.299 .300	.294 .294	.289 .290

TABLE 5.1 Moments of $W(\tau)$ (upper line) and of $\tilde{F}_\tau(w)$ (lower line) for $\tau = 0(.1).4$ and $n = 1, \ldots, 10$.

τ	σ	$\tilde{p}_o(t)$	p_1	α_τ	β_τ
0	.2353	.0064	.27	3.3905	1.8888
.1	.2559	.0119	.27	3.0675	2.0334
.2	.2747	.0194	.27	2.8000	2.1640
.3	.2917	.0298	.27	2.6076	2.3093
.4	.3069	.0431	.27	2.4814	2.4808

TABLE 5.2. The Parameters of the Mixed-Beta Distribution $F_\tau(w)$ for $\tau = 0(.1).4$. (σ denotes the standard deviations.)

n \ τ_i	0.0	0.1	0.2	0.3	0.4
0	0.0064	0.0119	0.0194	0.0298	0.0431
1	0.0065	0.0122	0.0203	0.0318	0.0466
2	0.0075	0.0147	0.0255	0.0411	0.0613
3	0.0104	0.0211	0.0375	0.0804	0.0894
4	0.0166	0.0332	0.0577	0.0906	0.1307
5	0.0273	0.0520	0.0870	0.1314	0.1836
6	0.0435	0.0785	0.1253	0.1819	0.2458
7	0.0662	0.1131	0.1723	0.2405	0.3145
8	0.0960	0.1557	0.2269	0.3052	0.3865
9	0.1333	0.2059	0.2879	0.3736	0.4585
10	0.1782	0.2628	0.3532	0.4430	0.5272
11	0.2303	0.3252	0.4209	0.5106	0.5894
12	0.2890	0.3914	0.4884	0.5735	0.6423
13	0.3531	0.4592	0.5529	0.6288	0.6836
14	0.4208	0.5261	0.6115	0.6736	0.7117
15	0.4902	0.5891	0.6613	0.7063	0.7265
16	0.5582	0.6448	0.6992	0.7249	1.0000
17	0.6213	0.6896	0.7226	1.0000	1.0000
18	0.8750	0.7183	1.0000	1.0000	1.0000
19	0.7139	1.0000	1.0000	1.0000	1.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000
Q_i	0.0704	0.0967	0.1273	0.1621	0.2000

TABLE 5.3. The CDF $\tilde{F}_{\tau_i}\left(\frac{\tau n}{L-\tau_i}\right)$, with $\tau = .1$, $\tau_i = 0(.1).4$, $L = s_U - s_L$; and the corresponding MGF Q_i.

8. References

[1] Chernoff, H. and Daly, J.F. (1957) "The Distribution of Shadows", *J. of Math. and Mechanics*, 6:567-584.

[2] Likhterov, Ya. M. and Gurin, L.S. (1966) "The probability of Interval Overlap by a System of Random Intervals", *Engineering Cybernetics*, 4:45-55 (in Russian)

[3] Yadin, M. and Zacks, S. (1982) "Random Coverage of a Circle with Applications to a Shadowing Problem", *J. App. Prob.*, 19:562-577

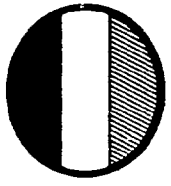
[4] Yadin, M. and Zacks, S. (1985) "The Visibility of Stationary and Moving Targets in the Plane Subject to a Poisson Field of Shadowing Elements", *J. App. Prob.*, 22:776-786

[5] Yadin, M. and Zacks, S. (1988) "Visibility Probabilities on Line Segments in Three-Dimensional Spaces Subject to Random Poisson Fields of Obscuring Spheres", *Naval Research Logistics*, 35:555-569

[6] S. Zacks and M. Yadin (1984) "Approximating The Probabilities of Detecting and of Hitting Targets and the Probability Distribution of the Number of Trials Along the Visible Portions of Curves in the Plane Subject to a Poisson Shadowing Process. Technical Report No. 3, Contract DAAG29-83-K-0176 with the U.S. Army Research Office.

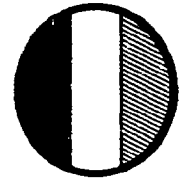
COMBAT MODELING

One afternoon of the 35th Conference on the Design of Experiments in Army Research, Development and Testing was devoted to a Special Session in the important area of combat modeling. First on the agenda was a paper by Donald H. McCoy entitled "Statistical Issues Related to Combat Modeling," and is published in these proceedings in the format of a slide presentation. The author advised the editor of these proceedings that most of the slides are self-explanatory; some are not. He figures that anyone who really wants to follow up would contact him. The title of the second paper planned for this session was "The Ballistic Research Laboratory Firepower Control Simulation from Inception to Validation," and is published in these proceedings. Unfortunately, its author, Ann E.M. Brodeen, was unable to attend the conference. Her place on the agenda was filled by a paper entitled "A Nonparametric Approach to the Validation of Stochastic Simulation Models" by William E. Baker and Malcolm S. Taylor. The last paper of the Special Session was presented by Eugene Dutoit. The attendees were given a thirty-page handout that he prepared for the convenience of the analyst who has to examine the results of force-on-force combat modeling. He provided these proceedings an abstract of this handout.



D. HUE McCOY

TRADOC ANALYSIS COMMAND

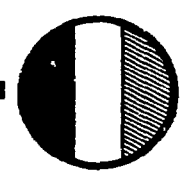


STATISTICAL ISSUES RELATED TO COMBAT MODELING

35th DESIGN OF EXPERIMENTS

OUTLINE

- PURPOSE
- TRAC MISSION
- BASIC MODELS (TOOL KITS)
- PROBLEM AREAS
- SPECIAL TOPICS
- ONGOING RESEARCH
- SUMMARY

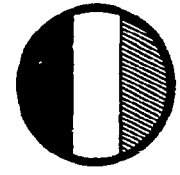


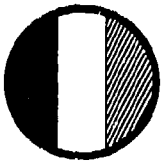
PURPOSE

OPEN DIALOGUE BETWEEN COMBAT MODELERS AND STATISTICAL EXPERTS

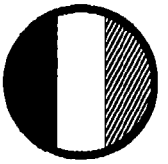
STIMULATE INTEREST ON YOUR PART

IMMEDIATE FEEDBACK FROM YOU





TRADOC ANALYSIS COMMAND (TRAC)



MISSION

The mission of TRAC is to conduct studies and analysis to support doctrine, combat and training developments in the Concept Based Requirements System; lead the TRADOC team conducting major studies and analysis; and develop and maintain analytic tools, scenarios and simulations for analysis and training of Airland Battle operations worldwide

GOALS

★ LEADERSHIP

A Command whose leaders at all levels possess the highest standards of ethics and professionalism, committed to excellence in mission accomplishment and the well-being of subordinates

★ CENTRALIZED COMMAND OF ANALYSIS

A Command which provides analytic service based on a well developed and managed study program with corporate development of taskers and plans and fully coordinated execution

★ INTEGRATED ANALYSIS

A Command whose analytic process ensures a balanced representation and linkage of the Army's functional areas and echelons in a worldwide joint/combined operations and environments which are simulated and analyzed

★ DIRECTED RESEARCH

A Command which continually explores emerging technologies and innovative approaches and harness them to improve the quality and timeliness of its analytic products

★ QUALITY PRODUCTS

A Command which is committed to excellance in Analysis and delivers timely, high quality analysis and simulations to meet the needs of Army leaders and trainers

★ PROFESSIONAL WORKFORCE

A Command composed of military and civilians who possess the highest ethical and professional standards, and the desire, skills and ability to produce the finest analyses for the army

PROBLEM SOLVING MODEL

1. PROBLEM

2. THREAT ANALYSIS

3. TECHNOLOGY ASSESSMENT

4. CONCEPTUAL ALTERNATIVES

5. OPERATIONAL AND ORGANIZATIONAL PLANS

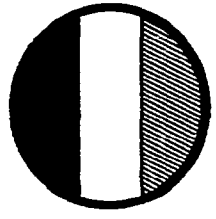
6. ANALYSIS

7. DECISION

8. IMPLEMENTATION

DOCTRINE
FORCE DESIGN
EQUIPMENT
LEADER DEVELOPMENT
TRAINING

TRADOC ANALYSIS COMMAND



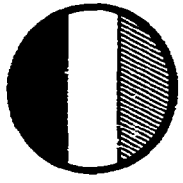
AAWS-M ANALYSIS

	PHS	PK/H	SSPK	RELOAD TIME	GUNNER EXPOSURE	PM INVESTMENT COST
TECHNOLOGY 1	0.9	0.6	0.5	30 SEC	15 SEC	$3.0 B
TECHNOLOGY 2	0.8	0.5	0.4	45 SEC	10 SEC	$2.5 B
TECHNOLOGY 3	0.8	0.4	0.3	70 SEC	20 SEC	$2.0 B

(ALL DATA REPRESENTATIVE)

GIVEN THESE CHARACTERISTICS, WHICH TECHNOLOGY SHOULD WE BUY?

TRADOC ANALYSIS COMMAND



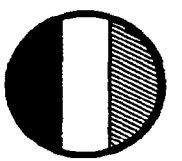
SAMPLE STUDIES

LHX - LIGHT HELICOPTER EXPERIMENTAL

HFM - HEAVY FORCE MODERNIZATION

FAADS - FORWARD AREA AIR DEFENSE STUDY

ARMOR/ANTI-ARMOR MIX ANALYSIS



Janus

USES
COEAs
TRAINING
CPX DRIVER
QUICK REACTION ANALYSES

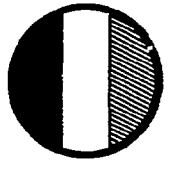
DESCRIPTION
STOCHASTIC
BATTALION-LEVEL
MAN-IN-THE-LOOP C^2
LITTLE AGGREGATION
DETAILED TERRAIN DATA
SMOKE/OBSCURATION PLAYED

HARDWARE
VAX W/VMS
TEKTRONIX
FORTRAN
-100K LINES

SETUP/EXECUTION
GRAPHICAL SCENARIO INPUT
ONE DAY TO TWO WEEKS
PER SCENARIO
WEAPON/TERRAIN DATA
-(5 MBYTES)
OUTPUT
-(5-25 MBYTES)
RUN TIME
-(3:1)
-(COMPUTER:REAL)

POST PROCESSING
GRAPHICAL PLAYBACK
FIRING EVENT SUMMARY
DETECTION EVENT SUMMARY
KILLER/VICTIM SCOREBOARDS
AMMO USAGE

TRADOC ANALYSIS COMMAND



CASTFOREM

USES
TRADE-OFF ANALYSES
O&O CONCEPT EVALUATIONS
PARAMETRIC ANALYSES
COEAs

DESCRIPTION
BATTALION-LEVEL
STOCHASTIC
EMBEDDED EXPERT C^2 SYSTEM
- -DECISION TABLES

REALISTIC BATTLEFIELD
- -TERRAIN, OBSCURATION

LASER WEAPONS

HARDWARE/SOFTWARE
VAX W/VMS
SUN, HP W/UNIX
RAMTEK
SIMSCRIPT II.5
- -140K LINES

SETUP/EXECUTION
4-6 MONTHS
- -NEW SCENARIO

2 WEEKS
- -MINOR MODS

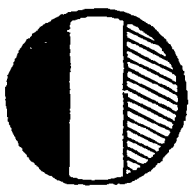
500+ DECISION TABLES
WEAPON/TERRAIN DATA, ETC.
- -(10-12 MBYTES)

OUTPUT
- -(8-10 MBYTES/REP)

RUN TIME
- -(2:1)
- -(COMPUTER:REAL)

POST PROCESSING
3-D GRAPHICAL PLAYBACK
COMPLETE HISTORY FILE
KILLER/VICTIM SCOREBOARDS
EXCHANGE RATIOS OVER TIME

TRADOC ANALYSIS COMMAND



VIC

USES

RESEARCH & EVALUATION
FORCE CAPABILITY
  REQUIREMENTS ANALYSES
COEAs

DESCRIPTION

CORPS-LEVEL
C ² W/DECISION TABLES
INTEL FUSION
AGGREGATED FORCES
CS/CSS
ELECTRONIC WARFARE
ENVIRONMENT
 -AGGREGATED STATISTICAL TERRAIN
 -SMOKE, WEATHER, ETC.

HARDWARE/SOFTWARE

VAX W/VMS
SUN W/UNIX
RAMTEK
SIMSCRIPT II.5
 -200K LINES

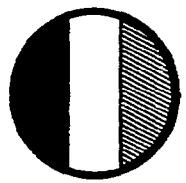
SETUP/EXECUTION

6-9 MONTHS
GRAPHICAL INPUT
INPUT
 -(10 MBYTES)
OUTPUT
 -(0.5 GIGA BYTES)
RUN TIME
 -(0.5:1)
 -(COMPUTER:REAL)

POST PROCESSING

GRAPHICAL PLAYBACK
KILLER/VICTIM SCOREBOARDS
UNIT STRENGTHS
FLOT TRACES
AMMO/POL USAGE
EXCHANGE RATIOS OVER TIME

TRADOC ANALYSIS COMMAND



AGGREGATION

DEFINITION

"LUMPED PARAMETER SYSTEM"

USES

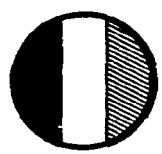
GROUPING OF SYSTEMS TO MAKE UNITS
-E.G., VIC IS BATTALION/REGIMENT RESOLUTION

PROBLEM AREAS

COMPLICATES LOS/DETECTION/ACQUISITION ALGORITHMS

COMPLICATES (INVALIDATES?) ATTRITION CALCULATIONS

NOTE: OFTEN REQUIRES THAT PROBABILITIES ARE USED AS MULTIPLIERS TO CREATE FRACTIONAL UNITS



DEAGGREGATION

DEFINITION

"UNDO COMMAND ON A LUMPED PARAMETER SYSTEM"

USE

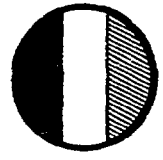
EXTRACT DETAILED INFO FROM THE SIMULATION

PROBLEMS

HOW TO DISTRIBUTE THE KILLS FROM MULTIPLE SYSTEMS FIRING AT THE SAME TARGET?

HOW TO DISTRIBUTE EXPENDED RESOURCES?

IN GENERAL, HOW TO SHOW CAUSE & EFFECT?



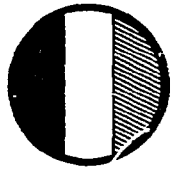
VIC ATTRITION

$$\Delta M_j = -\sum_i A_{ji} \; n_i \; \Delta t$$

ΔM_j = decrease in numerical strength of target group j

A_{ji} = attrition coefficient

n_i = current numerical strength of firing group i



ATTRITION COEFFICIENTS

INTERACTION OF FOUR PROCESSES

- LINE OF SIGHT
- TARGET ACQUISITION
- TARGET SELECTION
- FIRING AND KILLING

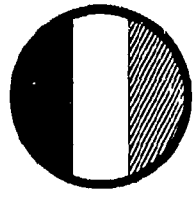
$$A = h \times \frac{1}{EFK} \times PF$$

WHERE

h = PROBABILITY THAT A TARGET BEING FIRED ON OR ACQUIRED WILL BE DESTROYED BY THAT FIRER BEFORE LINE OF SIGHT IS LOST OR THE TARGET IS DESTROYED BY ANOTHER FIRER.

EFK = EXPECTED TIME THAT A FIRER SPENDS FIRING AT A TARGET WHICH HE HAS ACQUIRED AND SELECTED WHEN THE ENGAGEMENT ENDS IN A KILL BY THE FIRER (CONDITIONAL KILL RATE).

PF = UNCONDITIONAL PROBABILITY OF FIRING

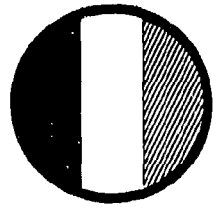


ATTRITION COEFFICIENT ASSUMPTIONS

EXPONENTIAL DISTRIBUTION OF

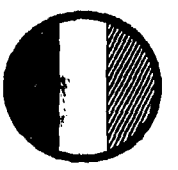
- TIME TO ACQUIRE
- DURATIONS IN VISIBLE OR INVISIBLE STATES
- TIME TO KILL

EFFECTS OF AN AGGREGATE GROUP CAN BE REPRESENTED BY A NUMBER OF "AVERAGE" ELEMENTS



BASIC PARAMETERS

- NUMBER OF FIRERS
- WEAPON CHARACTERISTICS
  - RANGE
  - FIELD OF REGARD
- NUMBER OF TARGETS IN RANGE
- PROBABILITY OF LINE OF SIGHT
- ACQUISITION RATE
- RATE OF MOVING OUT OF LINE OF SIGHT
- RATE THAT OTHER WEAPONS KILL TARGETS
- SELECTION PRIORITIES
- KILL RATE
- FIRING RATE



COMPARISON OF STOCHASTIC MODEL RESULTS

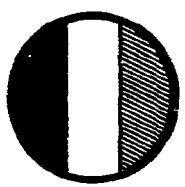
Case i	Case j
$LER_i = \frac{R_{Li}}{B_{Li}}$	$LER_j = \frac{R_{Lj}}{B_{Lj}}$

LER = LOSS EXCHANGE RATIO

R_L = RED LOSSES
B_L = BLUE LOSSES

STATISTIC

WILCOXON RANK SUM (MANN-WHITNEY) TEST
TIES ARE CONSIDERED
IS THERE A BETTER WAY ?



HOW MANY REPS FOR A STOCHASTIC MODEL ?

THE CURRENT METHOD :

-ASSUMES NORMAL DISTRIBUTION FOR i MEASURES

n^* = CURRENT NUMBER OF REPS

$$n_i = \left(\frac{Z * S_{x\,i}}{.05 * \bar{X}_i} \right)^2$$

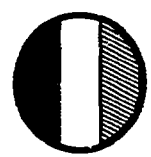
$n_{max} = MAX \{n_1, n_2 ..., n_i\}$

IF $n_{max} < n^*$ STOP

ELSE RUN $(n_{max} - n^*)$ MORE REPS !

IS THERE A BETTER WAY ?

IS THERE A DISTRIBUTION FREE METHOD?



RUN MATRIX DESIGN

PROBLEM

HOW TO DESIGN AN EXPERIMENT WITH THE MODEL?

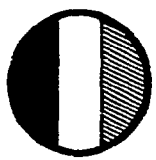
E.G., CONSIDER 6 MAJOR ALTERNATIVES

ALL COMBINATIONS 64 CASES

CPU TIME = 64 x 20 HOURS = 1280 HOURS

WHAT IS THE BEST WAY TO APPROACH THE DESIGN?

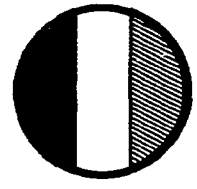
CAN WE USE CURRENT STUDY RESULTS TO BACK INTO A BETTER APPROACH?



SENSITIVITY ANALYSIS

- HOW SENSITIVE ARE MODEL RESULTS TO THE FORCE BALANCE BETWEEN COMBATANTS?
- HOW SENSITIVE ARE RESULTS TO CHANGES IN WEAPON PERFORMANCE MEASURES?
- HOW MANY SCENARIOS SHOULD BE PLAYED?

ALL THE ABOVE HAVE IMPLICATIONS FOR CAREFUL EXPERIMENTAL DESIGN.



MEASURES OF EFFECTIVENESS

RL - BL OR RL / BL

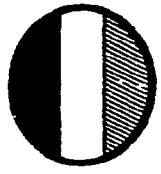
(RL/RO) - (BL/BO) OR (RL/RO) / (BL/BO)

RL = RED LOSSES
BL = BLUE LOSSES
RO = RED INITIAL FORCE
BO = BLUE INITIAL FORCE

GIVEN A MODEL AND A SITUATION, WHICH MOE ARE MOST SENSITIVE TO WEAPON PERFORMANCE?

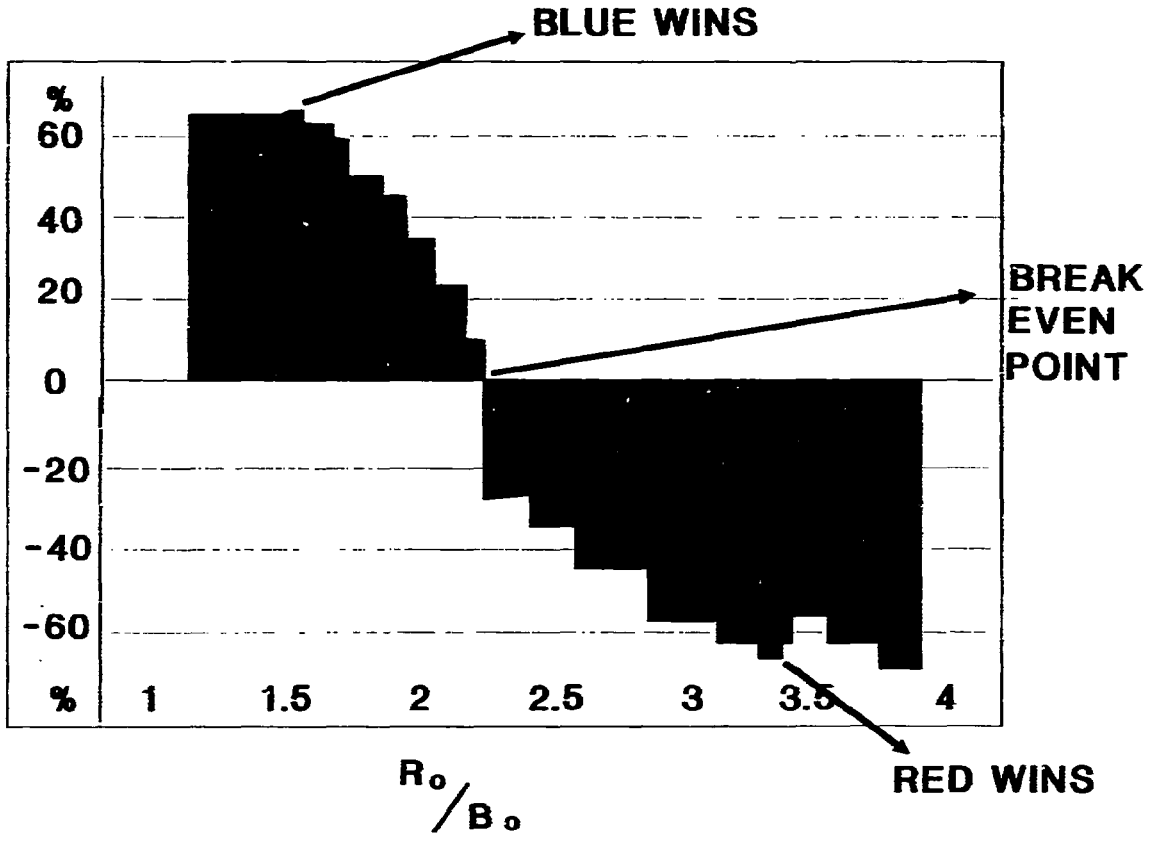
HOW DO CHANGES IN SITUATION AFFECT THE MOE?

WHICH MOEs ARE EASIEST TO DEAL WITH IN A STATISTICAL SENSE?



NORMALIZED LOSS DIFFERENCES

$$\frac{R_L}{R_o} - \frac{B_L}{B_o}$$



R_L - RED LOSSES

B_L - BLUE LOSSES

R_o - RED Starting Force

B_o - BLUE Starting Force

MODEL-TEST-MODEL

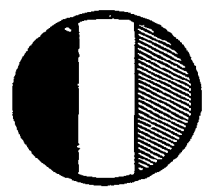
STEP 1: USE THE COMBAT MODEL TO SUPPORT TEST DESIGN

- -TRIAL DESIGN
- -CHECK EXECUTION TIMING
- -PLAN LOCATION OF TEST SUPPORT EQUIPMENT

STEP 2: EXECUTE THE TEST

STEP 3: USE THE MODEL TO EVALUATE THE TEST RESULTS

- -COMPARE TRIAL(SINGLE REP) W/ MULTIPLE REPS
- -EXTRAPOLATE TEST RESULTS
- -SENSITIVITY ANALYSES



SUPPRESSION

DEFINITION (SUPPRESS)

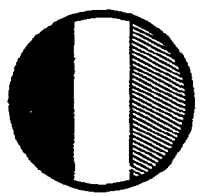
- TO PUT DOWN BY AUTHORITY OR FORCE
- TO RESTRAIN FROM A USUAL COURSE OF ACTION

PROBLEM

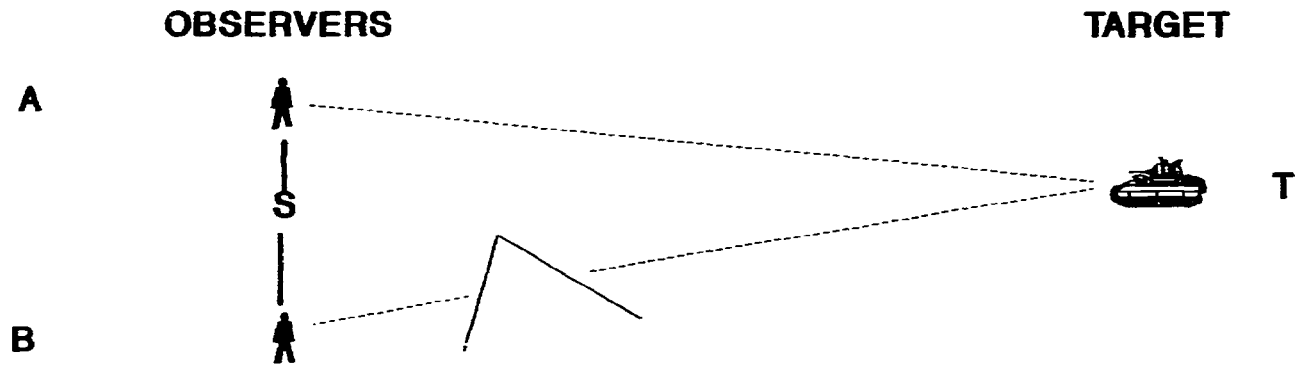
WHAT SHOULD SUPPRESSION BE BASED ON ?

- DANGER STATE ?
- VOLUME OF INCOMING FIRE ?
- PERCEIVED PROTECTION ?

HOW DO YOU TEST? OR MEASURE ?
HOW DO YOU MODEL ?



CORRELATED LOS



S = SEPARATION BETWEEN OBSERVERS

IF A SEES T DOES B ?
AS S⟶0 THEN A AND B SEE THE SAME ?

IF LOS IS AN EXPLICIT PROCESS THIS IS EASY !

HOW DO I MODEL THIS PROCESS IN
A LUMPED PARAMETER SYSTEM ?

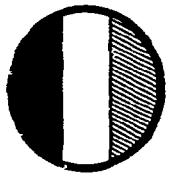
CALIBRATION OF LOW-RES MODELS WITH HI-RES MODELS

PROCEDURE

- DECOMPOSE LARGE SCALE LOW-RES MODELS INTO MANY SUB-BATTLES.
- PLAY SUB-BATTLES WITHIN HI-RES MODELS

QUESTIONS

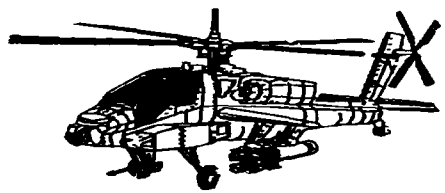
- THE MEANS MATCH! SO WHAT?
- GIVEN ALL THE DEGREES OF FREEDOM DOES CALIBRATION MAKE SENSE?



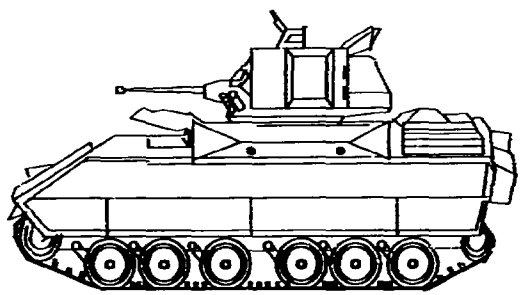
US ARMY

TRAINING AND DOCTRINE COMMAND

- TRAC -

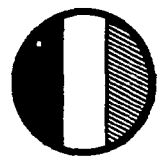


ARMOR ANTI-ARMOR MIX STUDY

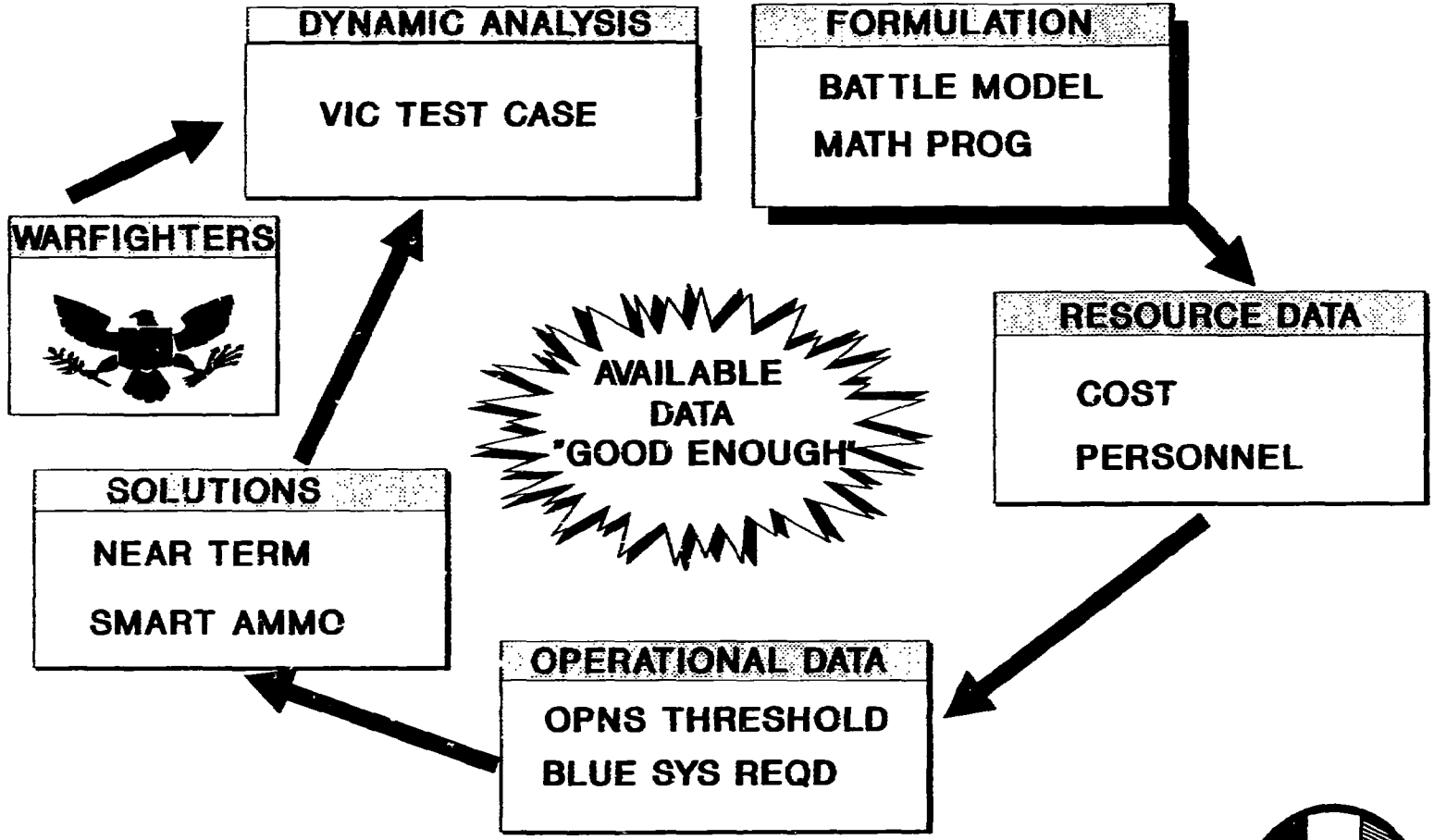


BASIC APPROACH

- DERIVE SYSTEM PERFORMANCE MEASURES FROM EXISTING SIMULATIONS
- USE MATH PROGRAMMING TECHNIQUES
- EXTRAPOLATE CORPS SOLUTION TO THEATER OPERATIONS
- DEVELOP SIMPLIFYING MODELS OF CORPS BATTLEFIELDS



METHODOLOGY



TRADOC ANALYSIS COMMAND

A3 SYSTEM MIXES

ECONOMIC MIX

OBJ FUNCTION:
MINIMIZE (SYSTEM & AMMO "COST")

RED PLANNING

SUBJECT TO :

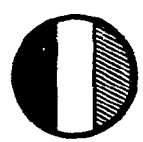
(1) CoFM REQUIREMENTS

(2) MIN % RED SYS TO KILL IN SS
(3) MIN % KILLS BY ZONE
(4) MAX % BATTLE KILLS

BLUE OPS

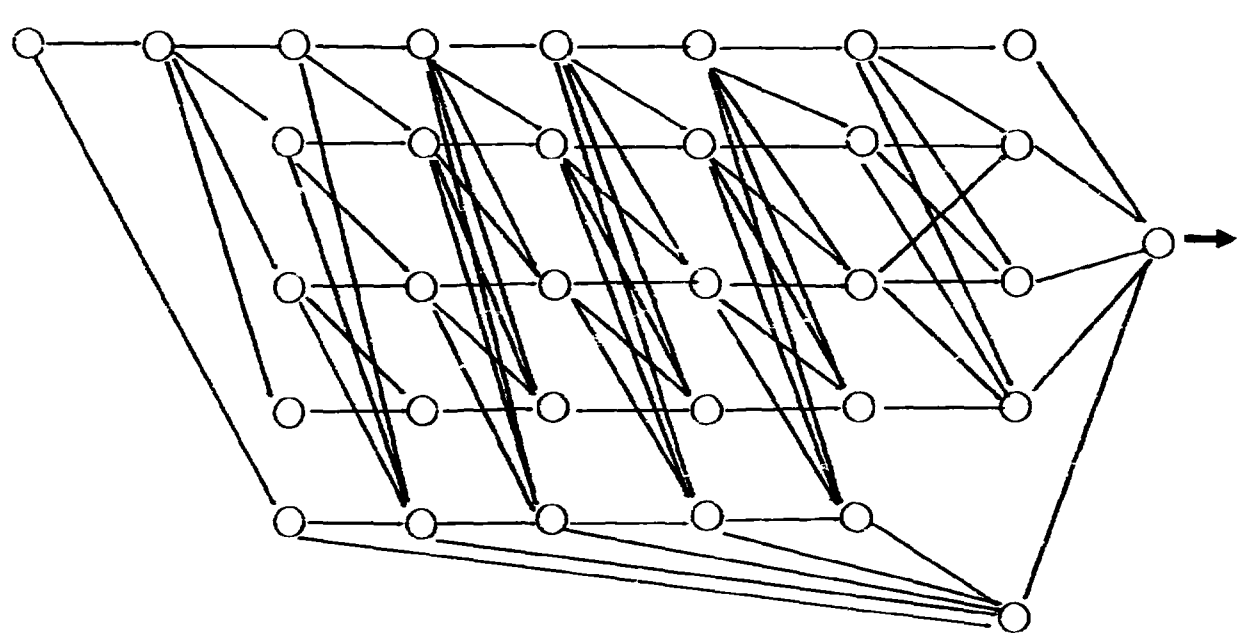
(5) PERSONNEL CEILING
(6) MIN/MAX BLUE SYSTEMS
(7) AMMUNITION CEILING

RESOURCES



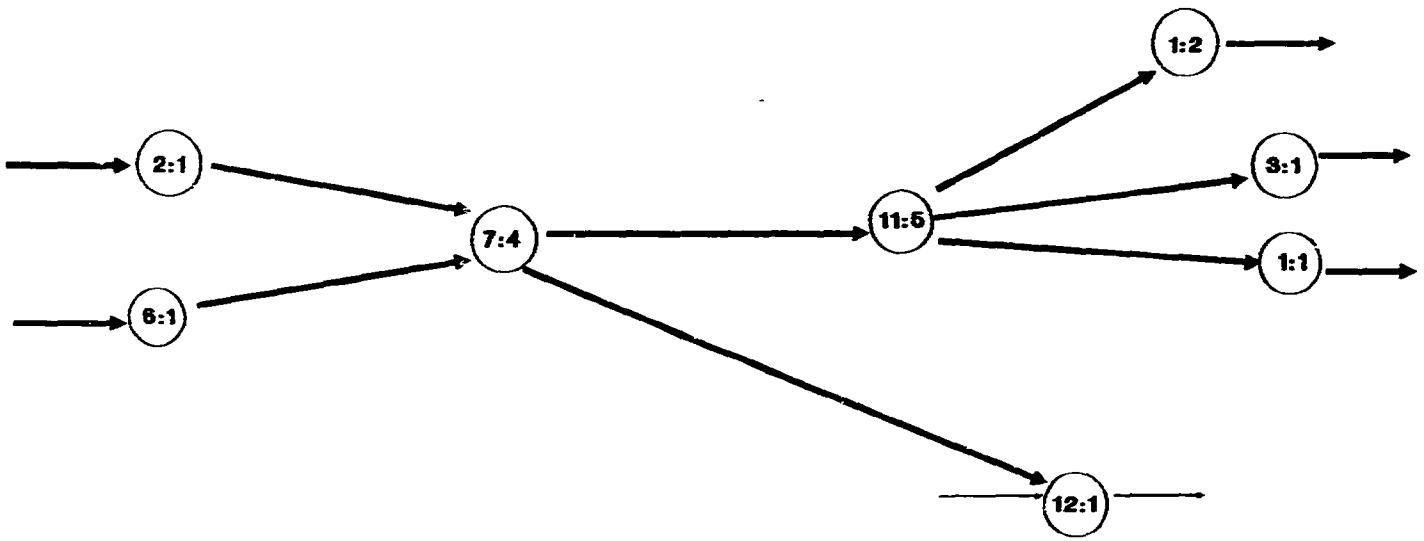
Stage-Transition Diagram

A/C STATUS \ STAGE	PRE-LAUNCH	LAUNCH	FLIGHT OVER FRIENDLY TERRITORY	PENE-TRATION	FLIGHT OVER HOSTILE TERRITORY	RELEASE	DAMAGE	RETURN
NO DEGRADATION								
LIGHT DEGRADATION								
SEVERE DEGRADATION								
BACK TO BASE								
LOSS								



TRADOC ANALYSIS COMMAND

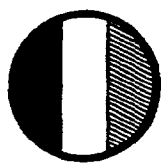
NETWORK OF MINIBATTLES



TRADOC ANALYSIS COMMAND

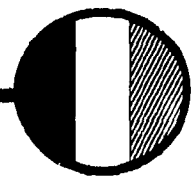
ANALYST'S WORK STATION

- GRAPHICAL PLAYBACK
- RELATIONAL DATABASE
- STATISTICAL PACKAGE
  - PARAMETRIC
  - NON-PARAMETRIC



CURRENT RESEARCH EFFORTS

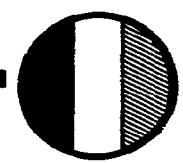
- AI BASED CORBAN REPLACEMENT (TRAC-FLVN/TRAC-LA)
- ATTRITION MODELING RESEARCH (ANCKER/GAFARIAN)
- ARMOR/ANTIARMOR WEAPONS MIX ANALYSIS
- GENERALIZED VALUE SYSTEM (Dr. PARRY)
- EXPERT SYSTEM FOR POST PROCESSING (TRAC-FLVN)
- JANUS/NTC COMPARISON (TRAC-MTRY)
- MODEL-TEST-MODEL W/CASTFOREM (TRAC-WSMR)
- TARGET SHADOWING (Dr. ZACKS)
- VIC EXPERT SYSTEM TO PRODUCE DECISION TABLES
- COMPETITIVE TRADEOFF MODELING (Dr. ROBINSON)
- HIERARCHICAL ANALYTICAL NETWORK SYSTEMS (Dr. CHARNES)



TRADOC ANALYSIS COMMAND

SUMMARY

- AGGREGATION/DEAGGREGATION
- ATTRITION MODELING
- MODEL CALIBRATION
- NUMBER OF REPLICATIONS
- COMPARISON OF MODEL RESULTS
- RUN MATRIX DESIGN
- SENSITIVITY ANALYSES
- MEASURES OF EFFECTIVENESS
- SIMPLIFIED ANALYSES



THE BALLISTIC RESEARCH LABORATORY FIREPOWER CONTROL SIMULATION FROM INCEPTION TO VALIDATION

Ann E.M. Brodeen

Director
U.S. Army Ballistic Research Laboratory
ATTN: SLCBR-SE-W
Aberdeen Proving Ground, MD 21005-5066
(301) 278-4109, AV 298-4109

Abstract

The Ballistic Research Laboratory Firepower Control Simulation (BRLFCS) is designed, in part, to support the on-going investigation of new ways of attacking the problem of data distribution on the battlefield. Ideally, prior to being utilized, the model should be validated, *i.e.*, tested whether or not it reasonably approximates the process of distributing tactical information across the battlefield. However, model validation generally assumes the availability of empirical data in order that some comparison may be made between the output generated by the model and real-world data. Unfortunately, a very limited empirical data base exists for the validation process. This paper provides an overview of BRLFCS related issues, *i.e.*, characteristics, supported applications, planned modifications. More importantly, a discussion of an approach proposed by Iman, Helton, and Campbell for validating large-scale computer models by replacing empirical data with model output will be presented in the context of the BRLFCS validation process [1,2].

I. Introduction

The BRLFCS is a large-scale information distribution model developed by the Weapon Systems Technology Branch (WSTB), System Engineering and Concepts Analysis Division (SECAD), BRL. Although a limited verification has been on-going as the model has evolved, the question has been continually raised as to whether the model could be statistically validated.

Currently, limited data exists for only a few tactical elements, *e.g.*, the fire support team headquarters (FIST HQ), the Field Artillery Battalion Fire Direction Center (FA Bn FDC), of the several included in the BRLFCS. This data was collected over the past several years from statistically designed firepower control experiments conducted in both research facility and field environments [3,4,5,6,7]. From the scope of the previous tests, it became evident that significant monetary and human resources must be expended to collect firepower control data for even a single tactical node. However, the WSTB is constructing its own Firepower Control Research Facility (FCRF) which should ease past resource burdens tremendously.

Statistical validation of the BRLFCS is beset by not only the lack of experimental data, but costly simulation runs and large numbers of input variables with differing characteristics, *e.g.*, qualitative and quantitative, discrete and continuous, ranges covering several orders of magnitude. These are all familiar problems facing anyone wishing to validate a large-scale simulation model. Although there has been innovative research done in this area, it, too, assumes the availability of at least some empirical data [8]. Fortunately, there is a technique which holds promise for validating large-scale models encumbered with the types of aforementioned problems. This generalized technique was proposed by Iman, Helton, and Campbell and is outlined in a two-part journal article [1,2].

This paper broadly outlines the techniques being proposed to validate the BRLFCS and the preliminary steps which have been completed at the time of the writing of this paper to place the validation process in motion. With this in mind, there are no results to report at this time. However, the author would like to solicit comments and critiques of the proposed solution to this problem, particularly from those who may have actually used the methodology.

II. The Ballistic Research Laboratory Firepower Control Simulation

a. Characteristics

The BRLFCS will be used to evaluate brigade (bde) area firepower control concepts for maneuver (mvr) and fire support elements. It is not intended for the model to be all-encompassing, but rather to provide an overview of the distribution of tactical information across the battlefield.

Some of the relevant features of the BRLFCS are presented in **Figure 1.** The version represented is a maneuver battalion (mvr bn) supported by field artillery units and is the version which will initially be validated. There also exists a brigade version which differs from the battalion version in scale only. Of particular importance with regard to the validation process is the fact the BRLFCS is a stochastic model, where stochastic model is hereby defined as

one in which, for each set of input values, a set of output values occurs with a certain probability. With such a model, any number of the input variables may be deterministic, so long as at least one is stochastic. Although a deterministic simulation was initially considered, in order to meet anticipated needs, a certain degree of randomness was built into the model, with the capability to suppress it if desired. Therefore, certain features of the BRLFCS were also designed to be stochastic. For instance, provision was built in to select the time a mission is initiated. These times may be either assigned explicitly, or the mission initiation rate, *i.e.*, number of missions per hour, can be given and the times assigned based on a random number string.

- Land Based
- Any mix of Blue Forces, mvr bde and below, including relevant fire support
- Supports any conflict for which data transmission requirements can be specified
- Resolution down to individual radios/data distribution units operates with ~260 in game; provision for 500
- Written in 'C'
- Input requirements: networks; units; transmission lengths and times; transmitter characteristics and locations; scenario data
- Outputs: unit and network loadings; queues; message and mission timelines
- Full scale runs made on a CRAY
  Reduced scale runs made on a Gould 9600
- Transmissions may be either TACFIRE or packet format
- Accomodates both TACFIRE and packet switching networks
- Processes performed in parallel

Figure 1. BRLFCS Features

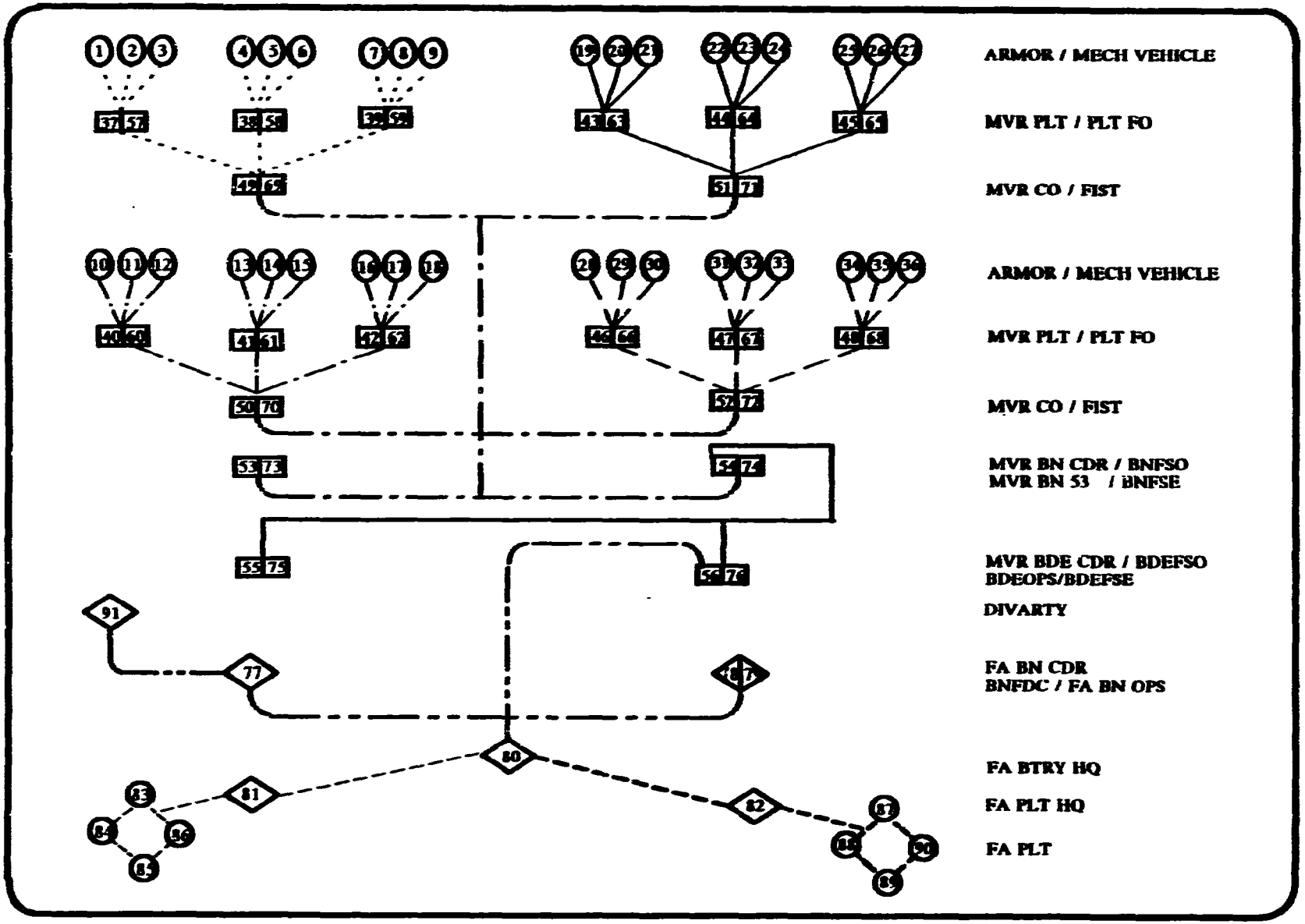


Figure 2. Units and nets for packet switching applications.

b. Concept

Although the simulation was planned so that it will be able to support future Army/DARPA Command & Control Project (ADDCP) activities, its principle function will be to demonstrate and evaluate the potential of new concepts of dynamic fire support management applications at the fighting level (bde and below), in particular the BRL Information Distribution System (IDS) fact-based technique [9].* In support of the IDS, the BRLFCS will be used to predict those links and/or procedures for data dissemination that result in excessive burdens on specific tactical nodes or networks, and to determine which aspects of the information flow have deleterious effects on mission duration time or asset utilization.

Overall, utilization of the computer simulation model should help narrow the focus of the on-going tactical computer science research, preventing it from pursuing "blind alleys".

c. Planned Modifications

Since the BRLFCS is designed to address specific issues while continuing to support the tactical computer science research effort, the simulation can be modified as needed. One such issue which may necessitate investigation, and which directly impacts the build up of queues in the network, is the manner in which high-priority missions entering a queue are handled. Normally this type of mission should be immediately advanced to the top of the queue for processing; however, the BRLFCS presently handles all missions on a first-in-first-out (FIFO) basis. While provision has already been built into the model to accomodate priority missions, the computer code has not yet been changed to address this issue.

Two other issues which the simulation does not presently address are unit attrition and multi-path information routings, *i.e.*, a more advanced scheme for routing packet message types (only) around the battlefield. These two issues are actually related in that, supposing a unit is operating at reduced efficiency, it may become desirable to reduce, or supress altogether, the amount of message traffic passing through that node. Under the existing routing algorithm pattern in the BRLFCS, this is impossible. As can be seen from **Figure 2**, the networks are now connected by single gateways (located at nodes 49 - 52, 54, 56, 78, and 80), thus forcing a transmitted packet message to follow a single path regardless of the number of times the message must be sent. Such a scheme may allow queues of unacceptable length to build up quickly.

* The basic concept of the IDS is to design a system capable of representing, storing, disseminating, and displaying facts in a tactical distributed computer environment.

III. Validation of the BRLFCS

a. Verification and "Face Validation"

During the course of its evolution, the BRLFCS has been undergoing almost continual verification; in other words, the correctness of the model is being established. This phase may be loosely described as "debugging" the program, *e.g.*, determining the reasonableness of values of certain model input variables and the correctness of the computer coding used. The 'C' program language allowed the BRLFCS to be easily structured into modules, or subprograms. By running the model using data employed in its construction, and observing the output from these modules, both the developer as well as "experts" knowledgeable about information distribution system models feel comfortable the model is behaving acceptably. When "experts" are insured a simulation is realistically representing the assumptions upon which it is based, this is often refered to as a model having "high-face validity".

Performing such a verification is allowing for a more efficient, simpler simulation design, which will eventually account for savings in computer time. Also, by previewing the output of the simulation modules, an experimenter is protected against anomolies which might occur in the responses when the model is used.

b. Anticipated Validation Approach(es)

It was originally envisioned that verification and "face validation" of the BRLFCS, as a complete system, would be the best that even recent advancements could offer, particularly in light of the difficulty in obtaining experimental data. Winter, *et al*, states, "The quality of the component models and the excellent knowledge of the random process along with a systematic verification must be a substitute for validation [10]."

However, a literature search unveiled a sensitivity approach to the validation of large-scale computer models, which to the author's knowledge, has not been utilized at the BRL. The approach is fully outlined in a two-part paper by Iman, Helton, and Campbell. Their approach focuses on the construction of a response surface as a replacement for the model. Underlying this approach is the substitution of model output for experimental data (due to the lack thereof). The remainder of this paper will highlight some of the features and strategies of this methodology which are being implemented into the validation of the BRLFCS.

Also planned is a statistical validation of the tactical nodes for which experimental data already exists (and which is independent of any data utilized in the development of the simulation). Referring to **Figure 2**, the tactical elements which will be validated are the FIST HQ, nodes 69 - 72; Field Artillery Battalion Commander (FA Bn Cdr), node 77; FA Battery Fire Direction Center (FA Btry FDC) positioned at the FA BTRY HQ, node 80. Although some similar type elements may be currently co-located with other types, or may even change their physical location in future applications, they are otherwise generic in nature, *e.g.*, the functions of FIST node 69 are equivalent to FIST node 70.

The approach for validating these nodes will entail a nonparametric procedure recently developed by Baker and Taylor for a stochastic computer simulation model [8].

IV. Strategies and Features

a. Preliminary Discussions of Model Input and Output Variables

Although numerous types of descriptive data will be collected during each simulation run, three model outputs have been identified as the measures that will be used in validating the BRLFCS. The three outputs are: 1) net usage, *i.e.*, the percent of time a specific net is occupied by message transmissions; 2) unit utilization, *i.e.*, the percent of time a specific unit is occupied with handling message traffic; and 3) mission duration.

The formats of the required BRLFCS inputs vary. Some require the simple assignment of a numerical value for program identification purposes only, *e.g.*, packet radios assigned a code of 6, while others are strictly deterministic or stochastic in nature. Still others may currently be designated as either deterministic or stochastic as mentioned in Section II.a.

Most of the present effort focuses on discussions being held between the model developer and the analyst. As a result of these discussions, several issues were identied as impacting the selection of an appropriate sensitivity technique. **First**, the developer has provided the analyst with an assessment of each input variable's anticipated impact on the model output based on his "expert" opinion. **Second**, for analysis purposes, it is being assumed that nonlinear relationships with the model outputs may exist. This does make the construction of an appropriate response surface a bit more tedious, but doable. However, it is also being assumed that there are no 2-way or above interactions among the input variables. **Third**, since the three output measures constitute a time dependent function of model input, each input variable must be examined to determine whether its importance changes significantly over time.

b. Input Vector and Significant Input Variables Selection Techniques

Obviously, in order to fit a response surface, model output must be obtained for various values of the input variables. The choice of which sampling scheme to use to select values for the input vectors presented a problem. Random sampling is not appropriate and, as for the other possibilities, *e.g.*, stratified sampling, double sampling, it nearly boiled down to a "grab bag" selection process. The sampling technique must take into consideration the possibility that one or more of the input variables might change in importance over time, as well as insure that all portions of each variable's sample space will be represented by input values, even when that distribution of values covers several orders of magnitude.

The Latin Hypercube Sampling (LHS) technique claims such advantages over other, more common, sampling schemes [1,2,12]. Another feature of this technique which makes it even more advantageous to the BRLFCS validation process, is that the probability distributions used with LHS do not necessarily have to be the "true" distributions In fact, if preferred, the range of values for the input variables may be used in place of probability distributions. For the majority of the BRLFCS input variables, their ranges of values is the only information available.

c. Input Variable Ranking and Response Surface Construction

One of the objectives of this sensitivity analysis will be to obtain a ranking of the potentially important input variables. This result will be used to help drive factors selected for future IDS testing. There are several regression techniques which may be used to select a "best subset" of the predictor variables. For the BRLFCS validation, stepwise regression will be utilized initially to construct a response surface based on a linear combination of the independent (input) variables [13].

Following an initial fit, several things should be checked, *e.g.*, is the fit adequate, consistency of independent variable selection if similar dependent variables are present, are the predictions reasonable. If the response surface is not providing a suitable representation for model output, then additional work is needed. Earlier it was mentioned that there is the possibility that the relationship between some, or all, of the BRLFCS input variables and the outputs is nonlinear. Iman, Helton, and Campbell suggest the use of rank regression as developed by Iman and Conover [14]. Rank regression is a relatively simple concept. Data are replaced with their corresponding ranks whereby usual regression procedures may be performed on these ranks.

d. Other Statistical Considerations

Only a few of the ideas that must be considered for the validation of the BRLFCS, or for that matter any sensitivity analysis, have been outlined using Iman, Helton, and Campbell as a guideline. No mention was made with regard to the actual validation of the response surface, the various diagnostic tools available for obtaining preliminary information for the construction of the surface, or data transformation. These issues are discussed in References [1,2].

V. Summary

The technique outlined by Iman, Helton, and Campbell appears to be a viable approach for validating the BRLFCS. Additionally, the use of the nonparametric technique developed by Baker and Taylor for stochastic models seems appropriate for performing a statistical validation of those tactical nodes for which experimental data exists.

A critique of these approaches, as well as suggested alternatives, are invited by the author.

References

[1] Iman, R.L., Helton, J.C., and Campbell, J.E., "An Approach to Sensitivity Analysis of Computer Models: Part I - Introduction, Input Variable Selection and Preliminary Variable Assessment", *Journal of Quality Technology*, Vol. 13, No. 3, July 1981.

[2] Iman, R.L., Helton, J.C., and Campbell, J.E., "An Approach to Sensitivity Analysis of Computer Models: Part II - Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis", *Journal of Quality Technology*, Vol. 13, No., 4, October 1981.

[3] Smith, J., Grynovicki, J., *et al*, "Fire Support Team Experiment", U.S. Army Ballistic Research Laboratory Memorandum Report, BRL-MR-3422, December 1984.

[4] McKaig, A.E., "Approach to Modeling the Distribution of the Message Service Times for the Fire Support Team Experiment", U.S. Army Ballistic Research Laboratory Memorandum Report, BRL-MR-3461, September 1985.

[5] Kaste, V.A., Brodeen, D.C., and Winner, W.A., "Description of the Digital Data Collected from FIST FDT&E II", U.S. Army Ballistic Research Laboratory Technical Report, BRL-TR-2676, September 1985.

[6] Grynovicki, J.O. and Smith, J.H., "Experimental Design and Analysis for the FIST Force Development Testing and Experimentation II", U.S. Army Ballistic Research Laboratory Memorandum Report, BRL-MR-3474, October 1985.

[7] Winner, W.A., Brodeen, A.E.M., and Smith, J.H., "Test Design and Analysis: Firepower Control Experiment, Part 12 of 12", U.S. Army Ballistic Research Laboratory Memorandum Report, BRL-MR-3612, June 1987.

[8] Baker, W.E. and Taylor, M.S., "A Nonparametric Statistical Approach to the Validation of Computer Simulation Models", U.S. Army Ballistic Research Laboratory Technical Report, BRL-TR-2696, November 1985.

[9] Smith, K.G., Heilman, E.G., and DiGiacinto, T.A., "Tactical Computer Science in Modern Battlefield Management", *Proceedings of the 1989 Symposium on Command and Control Research*, August 1989.

[10] Winter, E.M., Wisemiller, D.P., and Ujihara, J.K., "Verification and Validation of Engineering Simulations with Minimal Data", *Proceedings of the 1976 Summer Computer Simulation Conference*, 1976.

[11] Downs, A.R., "An Operator's Reference for the Ballistic Research Laboratory Firepower Control Simulation (BRLFCS)", U.S. Army Ballistic Research Laboratory Technical Report in preparation.

[12] McKay, M.D., Beckman, R.J., and Conover, W.J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", *Technometrics*, Vol. 21, No. 2, May 1979.

[13] Draper, N.R. and Smith, H., Applied Regression Analysis, Wiley & Sons, Inc., New York, 1981.

[14] Iman, R.L. and Conover, W.J., "The Use of the Rank Transformation in Regression", *Technometrics*, Vol. 21, No. 4, 1979.

A NONPARAMETRIC APPROACH TO THE VALIDATION OF STOCHASTIC SIMULATION MODELS

WILLIAM E. BAKER
MALCOLM S. TAYLOR

US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

ABSTRACT

For three decades interest in simulation modeling and simulation languages has been expanding, almost keeping pace with the phenomenal rate of growth of computer technology. Lagging somewhat behind has been attention to the validation of the resulting simulation models; that is, the establishment of some level of confidence that the model does, in fact, accurately mimic some real-world process. In the last fifteen years, research in validation techniques has been substantially increased; and one general conclusion has been that statistical tests are desirable in the validation process.

We have adapted a nonparametric statistical technique to validate a stochastic simulation, and this procedure has subsequently been applied to a computer model currently in use at the US Army Ballistic Research Laboratory. Monte-Carlo methods have provided an indication of the power of this statistical test.

KEYWORDS: Hypothesis Testing, Ranking Procedures, Power of Test

I. INTRODUCTION

For three decades interest in simulation modeling and simulation languages has been expanding, almost keeping pace with the phenomenal rate of growth of computer technology. Lagging somewhat behind has been the concern for the validation of the resulting simulation models; that is, the establishment of some level of confidence that the model does, in fact, accurately mimic some real-world process. In the last fifteen years, research in validation techniques has been substantially increased; and a consensus of general conclusions has formed:

1. validation is problem dependent - there is no one general validation technique, mainly because the output from a model may be independent or correlated, univariate or multivariate, stationary or dynamic, and so forth; in fact, the model itself may be deterministic or stochastic,

2. in general, absolute validity is nonexistent - once a particular technique has been established, the model is usually validated only for a specific purpose and over a specific range of values,

3. empirical data are necessary - in order to validate a model, some comparison of output data with real-world data must be made; furthermore, these empirical data must be independent of those used in construction of the model, and

4. statistical tests are desirable - of the many methods proposed for validating simulation models, the use of statistical tests seems to be preferred, possibly because of the ability to establish some level of confidence.

Nonparametric validation methods generally involve a procedure known as hypothesis testing. The initial step is to state a null hypothesis, usually "the simulation model is valid." Then a level of confidence is established, often 95%; and a particular test statistic is chosen. Two different errors are present in hypothesis testing. The first is called a Type I error and occurs when a true null hypothesis is rejected. If the level of confidence has been set at 95%, then it follows that the probability of a Type I error is 5%. However, in simulation model validation a Type II error is the more important to control; this occurs when a false null hypothesis is accepted. No level of confidence is pre-established to guard against accepting an invalid model; but, for any particular statistical test, a measure of the protection against this error is given by the power of the test, equal to the probability of rejecting the null hypothesis when it is false.

Unfortunately, there is a tradeoff between the two error types; as the level of confidence is increased (lower probability of a Type I error), the power of the test is decreased (higher probability of a Type II error). This implies that one way to increase the power of a test is to decrease the level of confidence in it. There are, however, more satisfactory ways; and they will be mentioned in the summary of this paper. The important point to remember is that when attempting to validate a simulation model using hypothesis testing, it is imperative that the statistical test be a powerful one.

II. LITERATURE REVIEW

As the electronic computer became a more powerful tool, computer simulation became a more viable method by which the behavior of a given process could be characterized. As early as the 1950's, articles were being published about computer modeling of entire systems; and soon after, specialized simulation languages were developed. The pioneers in this field realized the need for some assurance that the simulation output would be consistent with the empirical data that were available. However, prior to 1967 there was very little written that provided any explicit procedures which might be applied to determine the soundness of a computer model. In that year several papers concerning this problem were published, and two of them became a foundation upon which most subsequent efforts have been constructed.

In 1967, Fishman and Kiviat[1] provided definitions which differentiated the notions of verification and validation, terms which had previously been used interchangeably. "Verification determines whether a model with a particular mathematical structure and data base actually behaves as an experimenter assumes it does. Validation tests whether a simulation model reasonably approximates a real system." Most individuals working in this area today have subscribed to these definitions, although papers continue to be published which do not discriminate between the two ideas. Figure 1, taken from a paper by Winter, et. al.[2], is a Venn diagram illustrating the relationship between verification, validation, and other concepts within the field of computer simulation. Stone[3] believed the word assessment "... is preferable to validation which has a ring of excessive confidence about it." However, in this paper we will continue to consider validation as defined by Van Horn,[4] who expanded on the previous definition by giving it a somewhat statistical flavor. "Validation ... is the process of building an acceptable level of confidence that an inference about a simulated process is a correct or valid inference for the actual process."

---

[1] Fishman, G.S. and Kiviat, P.J., "Digital Computer Simulation: Statistical Considerations," Memorandum RM-5387-PR, The Rand Corporation, 1967.

[2] Winter, E.M., Wisemiller, D.P., and Ujihara, J.K., "Verification and Validation of Engineering Simulations with Minimal Data," Proceedings of the 1976 Summer Computer Simulation Conference, 1976.

[3] Stone, M., "Cross-Validating Choice and Assessment of Statistical Prediction," Journal of the Royal Statistical Society, Series B-36, 1974.

[4] Van Horn, R., "Validation," The Design of Computer Simulation Experiments, Duke University Press, 1969.

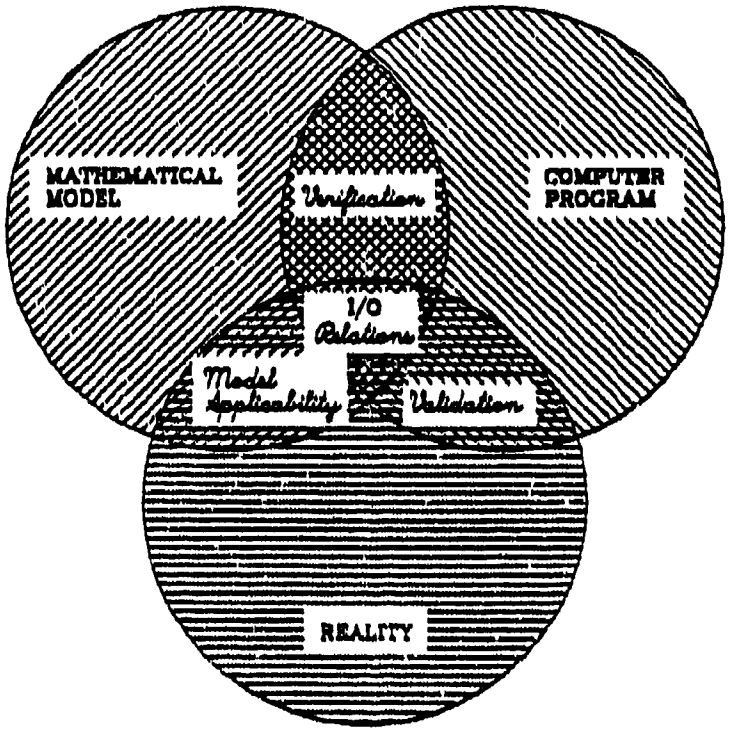


FIGURE 1: RELATIONSHIPS BETWEEN THE VARIOUS CONCEPTS OF A COMPUTER SIMULATION

The second influential paper to appear in 1967 was by Naylor and Finger.[5] In it they proposed a three-stage approach to validation of a computer simulation. This technique, or a modified version of it, has been used by numerous authors. Law[6] has augmented their approach with specific suggestions for each of the three stages:

1. develop high face-validity - insure that the simulation seems reasonable to those people who are knowledgeable in the area,

2. test the simulation assumptions - examine the data used in building the simulation and empirically test the assumptions drawn from those data, and

3. compare simulation output data with empirical data - use tests, statistical if possible, to determine a level of confidence in the simulation.

When attempting to validate existing models, the first two stages will often have already been completed by the developer of the simulation leaving only the third stage, potentially the most difficult.

---

[5] Naylor, T.H. and Finger, J.M., "Verification of Computer Simulation Models," Management Science, Vol.14 No.2, 1967.

[6] Law, A.M., Simulation Modeling and Analysis, University of Wisconsin, 1979.

Not everyone subscribes to the three-stage approach to validation. However, there does seem to be a general agreement that the third stage, comparing simulation output data with empirical data, is crucial. Sometimes obtaining empirical data in the region of applicability is very difficult, especially in engineering simulations. Winter, et. al.[2] mention in that case, "The quality of the component models and the excellent knowledge of the random process along with a systematic verification must be a substitute for validation." However, Fishman and Kiviat[4] are firm in their statement that " ... if no numerical data exist for an actual system, it is not possible to establish the quantitative congruence of a model with reality." In attempting to perform this third stage, Wright[7] suggests that three questions be considered:

1. how do we intelligently compare simulation output data with empirical data,
2. how do we collect and exploit the empirical data used in our tests, and
3. how do we transform the results of these tests into a confidence in the computer simulation?

Finally, Baird, et. al.[8] warn that the empirical data used for comparison with the simulation output data must be independent of those used in building the computer model; otherwise, we have only verification of the simulation.

Tytula[9] has divided the many methods used for the data comparison into five general categories:

1. judgemental comparison - this method seems to be the most widely used and includes graphical analysis and the comparison of common properties such as the mean and variance; it is easy to use and quite practical, but the impact of errors in judgement is difficult to assess,
2. hypothesis testing - this method includes goodness-of-fit tests, analysis-of-variance techniques, and nonparametric ranking methods; since this will be the category of interest in our report, the advantages and disadvantages will be discussed in the succeeding section,
3. spectral analysis - since the output of many simulation models is in the form of a time series, this method is particularly useful; however, it is difficult to relate the invalidity at a particular frequency to the overall simulation validity,

---

[7] Wright, K.D., "Validating Dynamic Models: An Evaluation of Tests of Predictive Power," Proceedings of the 1972 Summer Computer Simulation Conference, 1972.

[8] Baird, A.M., Goldman, R.B., Bryan, W.C., Holt, W.C., and Belrose, F.M., "Verification and Validation of RF-Environmental Models - Methodology Overview," Boeing Aerospace Company, 1980.

[9] Tytula, T.P., "A Method for Validating Missile System Simulation Models," Technical Report E-78-11, U.S. Army Missile Research and Development Command, 1978.

4. sensitivity analysis - this method can determine a range of parameter values and assumptions over which the simulation is valid, but it is usually difficult to analyze the effects of the characteristics drifting outside this range, and

5. indices of performance - this method is useful in ranking models; however, it is impossible to pick a value for a given index which will always imply a valid simulation.

Validation is a difficult process because, as Tytula[9] points out, no single satisfactory method exists. Most techniques are problem dependent; and, indeed, the output data of a simulation may be independent or correlated, univariate or multivariate, stationary or dynamic. In fact, Garrett[10] states that, "The critical dimension affecting the applicability of various techniques is that of the deterministic or stochastic nature of the output." Only a few authors have attempted to provide a general validation technique - see Gilmour[11] for an example. Most have developed methods which apply to a select subset of simulation models; and, even then, the simulation is often validated only for a particular purpose or over a particular range of values. In that case, care must be taken not to apply the simulation model outside the validated region.

III. VALIDATION PROCEDURES

In this paper we will be examining hypothesis testing as a method for validating stochastic computer simulation models. This type of procedure allows some level of confidence to be attached to the results. When employing hypothesis testing, several assumptions must usually be stated; but by using nonparametric ranking techniques we will eliminate one major (and often unjustifiable) assumption - that the data arise from a normal distribution.

Sargent[12] notes that for hypothesis testing we generally assume a null hypothesis that the simulation model is valid. Then by establishing a level of confidence for a particular statistical test, we fix the probability of a Type I error in which we reject a valid model. However, for simulation validation it is more important to minimize the probability of a Type II error, that is, accepting an invalid model. The magnitude of the Type II error can be determined by the power function of the statistical test where the power is the probability of rejecting a false null hypothesis. For a fixed sample size there is a tradeoff between the two error types, so that we can increase the power at the expense of the confidence level. Unfortunately, the power can not be computed against an alternative hypothesis as general as, "The simulation model is invalid"; and therefore, it must be examined against an array of different specific alternative hypotheses. Nevertheless, we continue to search for powerful

---

10 Garrett, M., "Statistical Validation of Simulation Models," Proceedings of the 1974 Summer Computer Simulation Conference, 1974.

11 Gilmour, P., "A General Validation Procedure for Computer Simulation Models," The Australian Computer Journal, Vol.5 No.3, 1973.

12 Sargent, R.G., "Developing Statistical and Cost-Risk Procedures for Validation of Simulation Models," U.S. Army Research Office Proposal Number 18201-M, 1980.

statistical tests with justifiable assumptions which will still provide acceptable levels of confidence.

Let $X = (x_1, x_2, ..., x_k)$ be a vector of inputs to a simulation model, and let y be an output resulting from X. Then y may take on many values is the case of a stochastic model. Let z be the corresponding value from the real-world process given the same input vector. In general, y will not be equal to z since X contains only a finite number of input variables; ostensively, the most relevant ones. The purpose of the simulation model is to mimic the real-world process. Thus, in attempting to validate it, we compare each empirical value with the corresponding model output generated under the same conditions; that is, the same values for the vector X.

Suppose there exist N pairs of data $(y_1, z_1), (y_2, z_2), . . ., (y_N, z_N)$ available for comparison, where each pair corresponds to a different input vector and where each y_i is itself be a vector of values from a stochastic model. Reynolds and Deaton[13] note that because each of the pairs was generated under different conditions, it would be incorrect to pool the data and proceed with the testing of our hypothesis. Rather, we must find a statistical procedure which examines each pair individually and then allows for the combination of these results into one overall test that provides reasonable power. With this as our goal, we propose to use a nonparametric statistical procedures - a process which combines independent cases of the Mann-Whitney test.

A stochastic model provides a set of output values that, for each given set of input values, occurs with a certain probability. Mihram[14] states that this "... probability ... serves as a measure of our human ignorance of the actual situation and its implications." Generally, the behavior of the system is too complicated to include all of the appropriate inputs in the computer model. Even if it were possible, the return in accuracy provided by such thoroughness may be small. Refinement of a computer model usually leads to stochastic modeling; and because of the abilities of today's computers, the use of such modeling has substantially increased.

Given M replications, output of the model becomes a set of values $y^1, y^2, ..., y^M$ for each set of input values which can be compared with (in our case) a single corresponding empirical value z. Recall that X is a vector of most, but not all, of the relevant input variables. Then z, given the value of X, is a random variable reflecting the random error due to the exclusion of certain factors from X. Also y, of course, is a random variable since the simulation model is stochastic. We would like to show that $F(y|X)$, the conditional distribution function of y, is equal to $G(z|X)$, the conditional distribution function of z for all $-\infty < y, z < \infty$ and for all X.

---

[13] Reynolds, M.R., and Deaton, M.L., "Comparisons of Some Tests for Validation of Stochastic Simulation Models," Commun. Statist. - Simula. Computa., Vol.11 No.6, 1982.

[14] Mihram, G.A., Simulation: Statistical Foundations and Methodology, Academic Press, Inc., 1972.

Considering N different input sets, the available data consist of N observations $(y_1^1, y_1^2, ..., y_1^M, z_1)$, $(y_2^1, y_2^2, ..., y_2^M, z_2)$, . . ., $(y_N^1, y_N^2, ..., y_N^M, z_N)$ of multivariate random variables, where the y^k's for any given observation share a common distribution. Mihram[14] suggests ranking $y_i^1, y_i^2, ..., y_i^M, z_i$ for each i; if the model is valid, we would expect the z_i to fall somewhere in the middle of such a ranking. This is the initial step in a procedure known as the Mann-Whitney test, a particular case in which one of the random variables, namely z_i, has a sample size of one. Since we are dealing with N observations, we need a method by which we can combine independent cases of the Mann-Whitney test; such a method has been proposed by Van Elteren[15] and referenced in a very clear example by Reynolds, et.al.,[16].

The Mann-Whitney test is a hypothesis test involving samples from two distributions that tests for equality of the distributions. For each input set X a sample of M output sets $y^1, y^2, ..., y^M$ is obtained from the computer simulation, and the empirical observation z provides another sample of size one. The following three assumptions are made:

1) both samples are random samples from their respective populations,

2) in addition to independence within each sample, there is mutual independence between the two samples, and

3) the measurement scale is at least ordinal.

The third assumption means that for any two observations on the random variable we can distinguish which is larger and which is smaller.

The null hypothesis is that $F(y|X) = G(z|X)$ for a given input set X. When we combine N of these tests, in the manner suggested by Van Elteren, we have the null hypothesis of $F(y|X) = G(z|X)$ for all $-\infty < y, z < \infty$ and for all X, which we can interpret as, "The simulation model is valid." Let R_i be the rank of z_i in the i^{th} observation $(y_i^1, y_i^2, ..., y_i^M, z_i)$; thus, R_i is an integer between 1 and M + 1. Then a test statistic T is defined as the sum of the R_i's over all N observations; that is, $T = \sum_i R_i$. Very high or very low values of T will cause rejection of the null hypothesis. The theory behind the Mann-Whitney test is given in Conover[17], and the combination of such tests is explained by Van Elteren[15].

A fourth assumption is usually made, that both samples consist of random variables from continuous distributions. This is to assure that there will be no zeros and, more importantly, no ties. However, for this test, a moderate number of ties is tolerable; and they are handled by assigning each of the tied values the average of the ranks normally due them.

---

[15] Van Elteren,P., "On the Combination of Independent Two Sample Tests of Wilcoxon," Bulletin de l'Institute International de Statistique, 37, 1960.

[16] Reynolds, M.R., Burkhart, H.E., and Daniels, R.F., "Procedures for Statistical Validation of Stochastic Simulation Models," Forest Science, Vol.27 No.2, 1981.

[17] Conover, W.J., Practical Nonparametric Statistics, John Wiley & Sons, Inc., 1971.

As mentioned earlier, a misuse of hypothesis testing as a method of simulation validation occurs when too little concern is shown for the power of the test. The power is the probability of rejecting an invalid model, and we would like this probability to be as close to one as possible. Unfortunately, the power can be calculated only for specific alternative hypotheses. In order to generate power curves for this combination of Mann-Whitney tests, it is convenient to make one additional, albeit restrictive, assumption; namely, the distribution of the y_i's is the same for each vector of input values, and similarly for the distribution of the z_i's. Although it would be preferable to avoid this assumption, it is necessary in order to test against specific alternative hypotheses - in this case, a shift in the mean.

Figure 2 shows some power curves for this test when the underlying distributions are normal and the mean of the distribution of the z_i's varies from zero. Recall that a true null hypothesis would indicate that the means of both F and G tend to be equal to zero . These curves were generated using a Monte-Carlo procedure which incorporated 10,000 replications. Note the increase in power as the number of observations increases. Figures 3-5 display some power curves for other alternative hypotheses, each figure assuming a different common distribution for F and G with a corresponding modification of the mean of G. Notice when the abscissa is equal to zero (when the null hypothesis is true), the probability of rejection is 0.05 - the value chosen for the probability of a Type I error. The faster the curve approaches one, the more powerful the test against that particular alternative hypothesis. Although very narrow in their scope, these results do provide us with an indication of the overall power of the test against a shift in location and allow us to determine the extent to which the probability of a Type II error might be reduced by an increase in sample size. Reynolds and Deaton[13] look at some test statistics similar to T designed to be more powerful against other alternative hypotheses.

IV. EXAMPLE

The Vulnerability Analysis for Surface Targets (VAST) model is a computer simulation currently in use at the Ballistic Research Laboratory to evaluate the effect of kinetic energy projectiles or shaped-charge threats against a single surface target.[18] It incorporates damage from both the primary penetrator and any associated spall fragments; but currently it is unable to handle damage resulting from blast, heat, and certain synergistic effects such as ricochets. Furthermore, there is a variety of opinions, estimates, and decisions, all based on the experience of the vulnerability analysts but generally providing vague and imprecise data, which subsequently serve as input to the simulation. Nevertheless, results demonstrate reasonable face validity, so an attempt at statistical validation of the model seems feasible.

A target description is produced by a separate computer code using a combination of geometric figures and, once generated, can be viewed from any orientation. After a viewing angle has been established, a rectangular grid is superimposed over the target in the plane orthogonal to that angle. From a (uniform) randomly-selected point within each grid cell, a

[18] Hafer, T.F. and Hafer, A.S., "Vulnerability Analysis for Surface Targets (VAST): An Internal Point-Burst Vulnerability Model," ARBRL-TR-02154, U.S. Army Ballistic Research Laboratory, 1979.

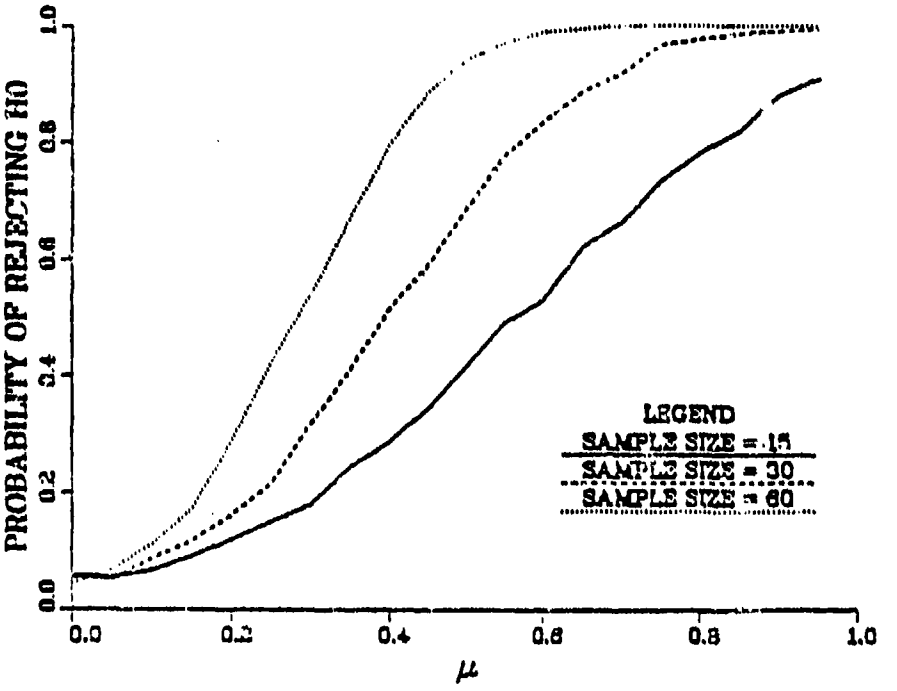


FIGURE 2: POWER OF 5%-LEVEL TEST
H0: F=G=NORMAL(0,1) VS. H1: F=NORMAL(0,1), G=NORMAL($\mu \neq 0$,1)

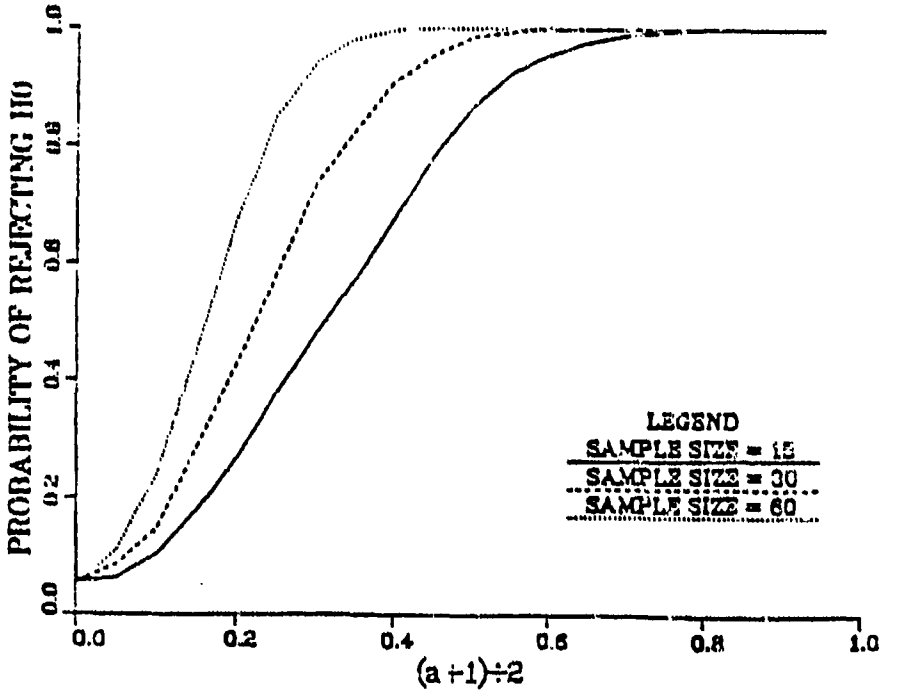


FIGURE 3: POWER OF 5%-LEVEL TEST
H0: F=G=UNIFORM(−1,1) VS. H1: F=UNIFORM(−1,1), G=UNIFORM($a \neq -1$,1)

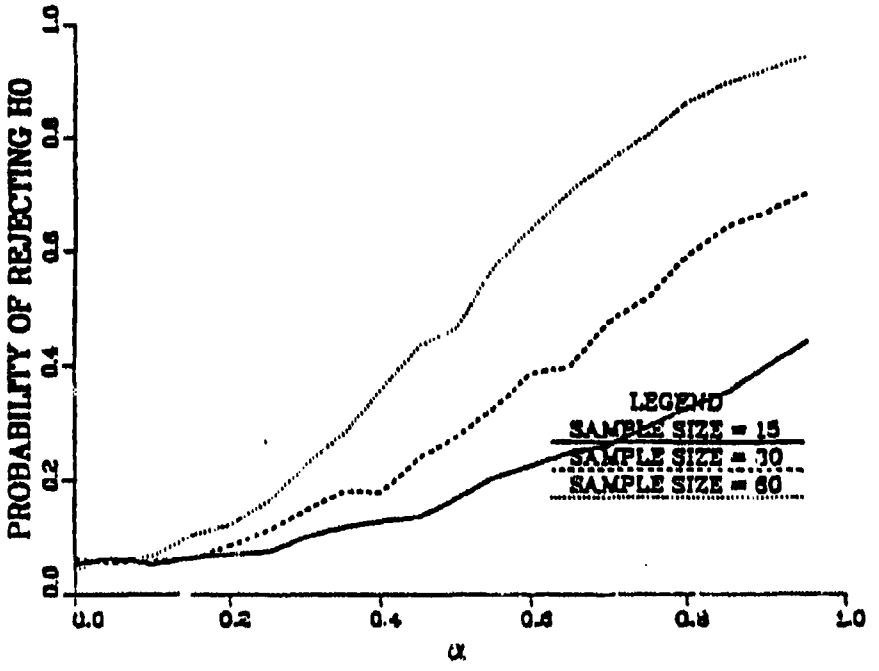


FIGURE 4 : POWER OF 5%-LEVEL TEST
H0: F=G=CAUCHY(0,1) VS. H1: F=CAUCHY(0,1), G=CAUCHY(α≠0,1)

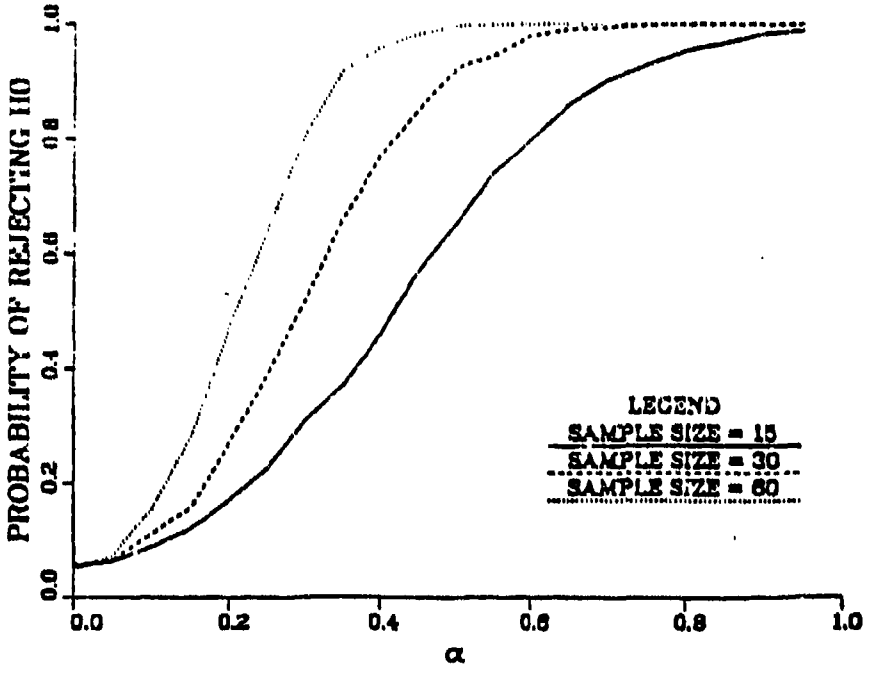


FIGURE 5: POWER OF 5%-LEVEL TEST
H0: F=G=LOGISTIC(0,1) VS. H1: F=LOGISTIC(0,1), G=LOGISTIC(α≠0,1)

ray is traced through the target; and a list is constructed of all components encountered. If a spall-producing component is encountered, spall rays are traced from that point of impact to all critical components in the target. These rays represent spall fragments whose size, shape, and velocity are chosen at random from specified distributions.

Along each individual ray, residual masses and velocities of the primary penetrator and associated spall fragments are used to calculate the probability of incapacitation for each critical component. These are then combined over all critical components and provide a loss of function (LOF) for the particular cell, further combined over all cells to provide a LOF for the particular orientation, and finally combined over several orientations to provide an overall LOF for the target.

Data were provided by vulnerability assessors who had estimated loss of function for a particular surface target based on their inspection of actual damage from a particular round of ammunition - in this case, the function evaluated was the mobility function. When attempting to compare model output with this empirical data, it was first necessary to determine the exact point of impact on the surface target during the live-fire exercise. Then the VAST model assumed that point of impact to be the origin of the ray representing the primary penetrator. Damage due to that ray and its associated spall rays were then combined to provide a LOF value which could be compared with the empirical datum point. Therefore, only one orientation was considered and, for that particular orientation, a ray originating at a specific point within only one cell was examined. Encountering a spall-producing component still required a random selection of spall characteristics; and because execution time was reduced, the model was run using thirty replications - the output data appear in Table 1. This output from the thirty two replications was compared with the empirical data, using the method proposed for stochastic simulations.

Table 2 contains the results. Recall that R_i is the rank of z_i in the i^{th} observation $(y_i^1, y_i^2, ..., y_i^M, z_i)$, and T is defined as the sum of the R_i's. Under the null hypothesis of a valid model, z_i has the same distribution as $y_i^1, y_i^2, ..., y_i^M$; and therefore, R_i is uniformly distributed over the values 1, 2, ..., M + 1. Lehmann[19] shows how to establish critical values against which the test statistic can be evaluated. Modifying his results by incorporating the number of tied observations, we can calculate the expectation of the test statistic,

$$E[T] = \frac{1}{2}[N(M+2)], \tag{1}$$

and the variance of the test statistic,

$$Var[T] = \frac{1}{12}[N M (M+2)] - \frac{1}{12[M+1]}\left[\sum_{i=1}^{N}\sum_{j=1}^{n_i}(d_{ij}^3 - d_{ij})\right], \tag{2}$$

where N is the number of observations, M is the number of replications of the model, and d_{ij} represents the number of tied values for the j^{th} tie in the i^{th} observation with n_i different ties in the i^{th} observation. Then $T^* = (T - E[T])/\sqrt{Var[T]}$ will have approximately a standard

[19] Lehmann, E.L., Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, Inc., 1975.

TABLE 1. LOSS OF FUNCTION VALUES - MOBILITY KILL

Shot Number	Replications														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
43	.7[illegible]	.7806	.6500	.8482	.6500	.7053	.7671	.6500	.6500	.6870	.8793	.6500	.8173	.7670	.7669
44	1.0[illegible]0	1.0000	1.0000	.1000	1.0000	1.0000	.1000	1.0000	1.0000	.1000	1.0000	.1000	.1000	1.0000	1.0000
45	1.0[illegible]	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
46	1.000[illegible]	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
47	.1434	.1000	.1148	.1713	.1000	.1169	.1320	.1000	.1482	.1000	.1000	.1305	.1185	.1000	.1000
[illegible]	.8380	.6500	.7879	.7069	.8408	.7804	.7671	.6911	.8755	.6500	.8575	.7113	.7440	.7122	.7670
[illegible]	.6500	.6500	.6500	.6500	.6849	.6500	.6500	.6500	.6500	.6906	.6500	.6500	.6500	.6903	.6500
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
51	.8733	.6315	1.0000	.7989	.9205	1.0000	.9548	.8020	.8058	.9772	1.0000	.6316	.9995	1.0000	.9996
52	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0100
53	.4005	.9990	.3902	.4725	.4002	.9991	.6031	.4003	.3913	.4003	.3910	.6243	1.0000	.4597	.4003
54	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
55	.8912	.9424	.8465	.9046	.6870	.8679	.9065	.8789	.9166	.9648	.8884	.9184	.8988	.8156	.8839
56	.3000	.3937	.3000	.0134	.3049	0.0000	.3000	.3050	.3097	.3000	.3048	.3000	.3099	.3053	.3099
57	.6444	.2423	.1000	.6500	.1000	.6498	.6499	.7244	.7032	.6500	.6500	.7039	.6648	.6500	.6500
58	1.0000	.9989	1.0000	1.0000	.9998	.9866	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.6580	1.0000	1.0000
59	.0059	.0512	.0500	.0500	.0487	.0060	.0489	.0690	1.0000	.0500	.0227	.0500	.0500	1.0000	.0521
60	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
62	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
64	.1000	.4015	.1000	0.0000	.0999	.1000	.1000	.1000	.1000	.1000	.3350	.0921	.1000	.5308	.1000
65	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
66	.8326	.8002	.7185	.7902	.8319	.7161	.7533	.6646	.8425	.6971	.6715	.7674	.7586	.8826	.7528
67	.8200	1.0000	1.0000	.9200	1.0000	1.0000	1.0000	.8960	.9299	1.0000	1.0000	1.0000	.9530	.7581	1.0000
68	.8945	.5497	.9978	.8978	.8020	.7220	.1000	.7197	.5500	.6486	.8942	.8995	.6498	.8982	.8937
69	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
70	.9997	1.0000	1.0000	1.0000	1.0000	1.0000	.9918	.9999	.6987	.6500	1.0000	.9765	.8996	1.0000	1.0000
71	.1000	.1337	.1628	1.0000	1.0000	1.0000	1.0000	1.0000	.2551	1.0000	.1000	.1000	.1000	.1000	1.0000
72	.9859	.9979	.8519	.9542	.9918	.9740	.9877	.9538	.9929	.8442	.9561	.9968	.8705	1.0000	.9762
73	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
74	.6968	.4015	.8521	.4014	.4015	.6609	.6813	.8774	.4015	.6925	.8356	.6909	.6574	.6960	.4015
75	.6374	.6465	.6854	.6757	.6229	.5580	.6254	.6805	.8372	.2984	.6761	.6499	.7939	.6269	.7078
76	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 1. LOSS OF FUNCTION VALUES - MOBILITY KILL (Cont'd)

Shot Number	Replications															
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Average
43	.6500	.7543	.6500	.6722	.7725	.6500	.6500	.7546	.6500	.6666	.7639	.7508	.6500	.7670	.7665	.719
44	.1000	1.0000	1.0000	1.0000	.1000	.1000	1.0000	.1000	1.0000	.1000	1.0000	1.0000	1.0000	1.0000	1.0000	.700
45	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
46	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
47	.1149	.1000	.1000	.1000	.1181	.1180	.1181	.1000	.1000	.1000	.1000	.1370	.1507	.1328	.1181	.116
48	.9885	.7197	.9895	.9903	.7127	.7258	.7673	.7672	.6716	.6500	.7429	.7670	.9994	.7253	.6712	.776
49	.6498	.6500	.6500	.6500	.6500	.6500	.6500	.6814	.6500	.6500	.6500	.6500	.6500	.6500	.6500	.655
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
51	.9196	.4761	.8407	1.0000	.8305	.5897	1.0000	1.0000	.9490	.9998	.9506	.6395	.8274	.9979	.9998	.881
52	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.967
53	.4000	.6611	.4005	.3909	.3919	.3214	.3996	.4004	.4005	.3909	.9993	.4000	.4004	.4005	.3887	.503
54	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
55	.8651	.8547	.8494	.9874	.7480	.8917	.8577	.9414	.9282	.9841	.9127	.9084	.9137	.9421	.9063	.890
56	.3000	.3000	.3000	.3093	.3051	.3052	.3000	.3000	.3000	.3000	.3052	.3000	.3050	.3000	.3000	.286
57	.6500	.7922	.6500	.2388	.1000	.7039	.1000	.7582	.1971	.6500	.7076	.7424	.3326	.2046	.4268	.523
58	1.0000	1.0000	.9868	.9671	1.0000	1.0000	1.0000	.9976	1.0000	.9999	1.0000	.9967	.9992	1.0000	.9911	.986
59	1.0000	.0519	.0143	.0500	.0450	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.457
60	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
62	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.1900	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.973
64	.5123	.1000	.3998	.1000	.0948	.4015	.5301	.5498	.1000	.1000	.1000	.1000	.1000	.1000	.5499	.207
65	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
66	.7885	.5034	.7791	.8440	.7291	.7443	.6762	.7982	.6942	.7425	.7236	.5866	.7499	.6108	.5988	.735
67	1.0000	.8419	1.0000	1.0000	1.0000	.9935	1.0000	1.0000	1.0000	1.0000	.9893	1.0000	.9974	1.0000	1.0000	.970
68	.9993	.8993	.6560	.8998	.6511	.9000	.5500	.8980	.5500	.5500	.6611	.5500	.5500	.9985	.7013	.738
69	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
70	.9995	1.0000	1.0000	.9970	1.0000	.9999	1.0000	1.0000	.9975	.9928	.9121	.9838	.6500	1.0000	.7184	.949
71	1.0000	.1000	1.0000	.1000	.3262	1.0000	.1827	.1833	.1000	1.0000	.1000	1.0000	.1379	.1000	1.0000	.513
72	1.0000	.9826	.9948	.9150	.9942	.9919	.9937	.9994	.9876	.7080	.9999	.9763	.9958	.9969	.8713	.958
73	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
74	.8543	.7007	.6941	.7431	.4015	.4015	.4015	.4015	.9095	.8486	.4015	.4015	.6664	.6585	.4012	.608
75	.8733	.6474	.6182	.6344	.6707	.6490	.8091	.6772	.8318	.6339	.7693	.6497	.7537	.6812	.7571	.679
76	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000

TABLE 2. HYPOTHESIS TEST		
Shot Number	Empirical Value	Rank within Model Values
43	.734	16
44	.145	11
45	1.000	16
46	1.000	16
47	.100	8
48	.900	27
49	.930	31
50	1.000	16
51	.145	1
52	1.000	16
53	.668	27
54	1.000	16
55	1.000	31
56	.905	31
57	.550	11
58	1.000	22.5
59	1.000	24.5
60	.050	1
62	1.000	16.5
64	.100	13.5
65	1.000	16
66	.668	6
67	.953	7.5
68	1.000	31
69	1.000	16
70	1.000	24
71	1.000	24.5
72	1.000	30
73	1.000	16
74	.905	30
75	.668	15
76	1.000	16

$\sum$ Ranks = 584

Critical T-Values (α = 0.05) = 435 (lower), 589 (upper)

Critical T-Values (α = 0.10) = 447 (lower), 577 (upper)

normal distribution. For our example we have 32 observations, 30 replications, and 51 instances of tied values with varying numbers of ties; in this case E [T] = 512 and Var [T] = 1521. We can calculate critical values by evaluating the equation $T = 39\hat{z} + 512$, where $\hat{z}$ is the $\alpha/2$ percentile of the standard normal distribution. As shown at the bottom of Table 2, there is insufficient evidence to reject the null hypothesis at an α-level of 0.05; however, at an α-level of 0.10, the null hypothesis would be rejected.

Since the null hypothesis could not be rejected at an α-level of 0.05, we must be concerned with the possibility of a Type II error; that is, accepting an invalid model. Figures 2-5 demonstrate the power of these tests against an alternative consisting of a shift in the mean. Figure 3 shows that the power of this test is very good if F (the distribution of the model output) and G (the distribution of the empirical data) are both uniform. However, as seen in Figure 4, if F and G are both Cauchy, then the power of the test is rather poor.

Reynolds and Deaton[13] have proposed other test statistics more powerful against different alternatives; but for the loss of function data where empirical results that are close to the value one tend to be assigned that value, a shift in the mean seems to be an appropriate alternative hypothesis. Since the power against this particular alternative is fairly good overall, our confidence in the hypothesis tests tends to increase. However, we would like to be able to make these tests and other tests still more powerful and, in the future, will be exploring methods to accomplish this.

V. SUMMARY

When referring to computer simulation models, a few authors continue to use the words verification and validation interchangeably; however, most distinguish between the two terms. Verification of a computer model assures that the simulation is behaving as the modeler intends, while validation assures that the simulation is behaving as the real world does. Verification is the process of debugging a computer program; validation is making it consistent with reality.

Prior to 1967 very little was written concerning the validation of simulations; but much has appeared since then, and there has been general agreement on several points - the most important being that to validate a computer simulation model, empirical observations are necessary and statistical tests are desirable. All validation techniques can be placed into one of five categories: judgemental comparisons, hypothesis testing, spectral analysis, sensitivity analysis, and indices of performance.

Nonparametric ranking techniques are one class of statistical hypothesis tests. We have advocated a combination of independent Mann-Whitney tests as a validation procedure for stochastic simulation models. This is a statistical test which assesses empirical data to provide a certain level of confidence in the computer model. The main disadvantage is the same as that of all hypothesis testing techniques; namely, their concern for protecting against Type I errors, sometimes at the expense of Type II errors. A Type I error results in rejecting a valid simulation model - unfortunate, but not as potentially dangerous as accepting an invalid simulation model, which is known as a Type II error. For any particular test we can get an indication of the probability of a Type II error by generating a series of curves that will allow

us to examine the power of the test against various alternatives.

Power is defined as the probability of rejecting a false null hypothesis, and we would like this value to be as close to one as possible. For our advocated test we have evaluated the power for some specific alternative hypotheses by incorporating a Monte-Carlo procedure into a computer program, which allowed us to perform thousands of replications. Each replication represents a case in which the alternative hypothesis was true, and we determined whether or not the test rejected the null hypothesis. Obviously, we can not compute power against an alternative hypothesis as general as, "The simulation model is invalid." However, in being more specific we are forced to examine an array of different alternative hypotheses; and while a test may be powerful against a subset of these alternatives (such as a shift in the mean of a distribution), it might be less so against others. The most we can hope for is reasonable power against alternatives important to a particular investigation. The combination of independent Mann-Whitney tests appears to have reasonable power against a shift in the mean, but we would like to be able to increase it.

For any given alternative hypothesis there are several ways of increasing the power. One such way can be seen in Figures 2-5 - increasing the number of observations. Another way is to reduce the level of confidence in the test itself; that is, allow the probability of a Type I error to increase. Because of the importance in this area of computer simulation validation, we hope to develop other ways to make these tests more powerful against a wide range of alternatives while still permitting them to provide acceptable levels of confidence in their results.

REFERENCES

1. Fishman, G.S. and Kiviat, P.J., "Digital Computer Simulation: Statistical Considerations," Memorandum RM-5387-PR, The Rand Corporation, 1967.

2. Winter, E.M., Wisemiller, D.P., and Ujihara, J.K., "Verification and Validation of Engineering Simulations with Minimal Data," Proceedings of the 1976 Summer Computer Simulation Conference, 1976.

3. Stone, M., "Cross-Validating Choice and Assessment of Statistical Prediction," Journal of the Royal Statistical Society, Series B-36, 1974.

4. Van Horn, R., "Validation," The Design of Computer Simulation Experiments, Duke University Press, 1969.

5. Naylor, T.H. and Finger, J.M., "Verification of Computer Simulation Models," Management Science, Vol. 14 No. 2, 1967.

6. Law, A. M., Simulation Modeling and Analysis, University of Wisconsin, 1979.

7. Wright, R.D., "Validating Dynamic Models: An Evaluation of Tests of Predictive Power," Proceedings of the 1972 Summer Computer Simulation Conference, 1972.

8. Baird, A.M., Goldman, R.B., Bryan, W.C., Holt, W.C., and Belrose, F.M., "Verification and Validation of RF-Environmental Models - Methodology Overview," Boeing Aerospace Company, 1980.

9. Tytula, T.P., "A Method for Validating Missile System Simulation Models," Technical Report E-78-11, U.S. Army Missile Research and Development Command, 1978.

10. Garrett, M., "Statistical Validation of Simulation Models," Proceedings of the 1974 Summer Computer Simulation Conference, 1974.

11. Gilmour, P., "A General Validation Procedure for Computer Simulation Models," The Austrailian Computer Journal, Vol. 5 No. 3, 1973.

12. Sargent, R.G., "Developing Statistical and Cost-Risk Procedures for Validation of Simulation Models," US. Army Research Office Proposal Number 18201-M, 1980.

13. Reynolds, M.R. and Deaton, M.L., "Comparisons of Some Tests for Validation of Stochastic Simulation Models," Commun. Statist. - Simula. Computa., Vol. 11 No. 6, 1982.

14. Mihram, G.A., Simulation: Statistical Foundations and Methodology, Academic Press, Inc., 1972.

15. Van Elteren, P., "On the Combination of Independent Two Sample Tests of Wilcoxon," Bulletin de l'institute International de Statistique, 37, 1960.

16. Reynolds, M.R., Burkhart, H.E., and Daniels, R.F., "Procedures for Statistical Validation of Stochastic Simulation Models," Forest Science, Vol. 27 No. 2, 1981.

17. Conover, W.J., Practical Nonparametric Statistics, John Wiley & Sons, Inc. 1971.

18. Hafer, T.F. and Hafer, A.S., "Vulnerability Analysis for Surface Targets (VAST): An Internal Point-Burst Vulnerability Model," ARBRL-TR-02154, U.S. Army Ballistic Research Laboratory, 1979.

19. Lehmann, E.L., Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, Inc., 1975.

SMALL SAMPLE TESTS IN SUPPORT OF COMBAT MODELING

EUGENE DUTOIT
U.S. ARMY INFANTRY SCHOOL
FORT BENNING, GEORGIA

ABSTRACT: THIS HANDOUT/REPORT HAS BEEN PREPARED FOR THE CONVENIENCE OF THE ANALYST WHO HAS TO EXAMINE THE RESULTS OF FORCE-ON FORCE COMBAT MODELING. THESE TESTS HAVE EXACT QUANTILE DECISION CRITERIA FOR SMALL SAMPLE DATA SETS. THIS IS THE USUAL SITUATION FOR LARGE, COMPLEX,MANPOWER RESOURCE INTENSIVE AND TIME CONSUMING FORCE-ON-FORCE MODELS. HOPEFULLY THIS APPLICATION PAPER WILL PROVIDE A REFERANCE THAT WILL GIVE SOME OF THE COMMON (AND EVENTUALLY THE UNCOMMON) STATISTICAL DECISION CRITERIA APPROPRIATE FOR SMALL SAMPLES. POST-HOC/MULTIPLE COMPARISON TECHNIQUES WILL BE PROVIDED WHERE AVAILABLE. THE HANDOUT IS LENGTHY (30 PAGES) AND NOT APPROPRIATE FOR PUBLICATION IN THE CONFERENCE PROCEEDINGS. IF YOU WANT A COPY OF THE HANDOUT PLEASE CALL OR WRITE:

GENE DUTOIT
AV 835-3165/3166
COMM (404) 545-3165/3166

COMMANDANT
U.S. ARMY INFANTRY SCHOOL
ATTN: ATSH-CD-CS-OR (DUTOIT)
FORT BENNING, GA 31905-5400

REQUEST THE USERS OF THIS PRELIMINARY COMPILATION PROVIDE FEEDBACK. ARE THE RANGES OF THE VARIABLES (TREATMENTS,REPLICATIONS,BLOCKS, ETC. SUFFICIENT TO MEET YOUR NEEDS? WHAT OTHER TESTS AND PROCEDURES WOULD YOU LIKE TO ADD TO THIS INITIAL LIST?

TABLE OF CONTENTS:

2. TWO-WAY - (CORRELATED SAMPLES) [K = TREATMENTS, N = BLOCKS]

A. 2 CORRELATED SAMPLES [N = 4, 5, ..., 7], [α = .05, .10, .20]

B. K SAMPLES, FRIEDMAN [K = 3, 4, 5], [N = 3, 4, ..., 8], [α = .05, .10, .15]

B1. POST-HOC, ALL COMPARISONS [K = 3, 4, ..., 6], [N = 2, 3, ..., 8], [α = .05]

B2. POST-HOC, TREATMENT VS CONTROL, ONE-SIDED [K = 3 COMPARISONS], [N = 2, 3, ..., 7], [α = .05, .10, .15]

B3. POST-HOC, TREATMENT VS CONTROL, TWO-SIDED [K = 3 COMPARISONS], [N = 2, 3, ..., 7], [α = .05, .10, .15]

B4. SEVERAL OBSERVATIONS FOR EACH TREATMENT IN EACH BLOCK - NO TABLES PROVIDED BUT INCLUDED FOR ITS UTILITY

C. ORDERED ALTERNATIVES [K = 3, 4, ..., 6], [N = 2, 3, ..., 7], [α = .05]

3. TESTS FOR DISTRIBUTIONS [K = TREATMENTS, N = REPLICATIONS]

A. TWO DISTRIBUTIONS [SMIRNOV], [N = 3, 4, ..., 10], [α = .05, .10, .20]

B. THREE DISTRIBUTIONS [BIRNBAUM-HALL], [N = 4, 5, ..., 10], [α = .05, .10, .20]

C. K DISTRIBUTIONS [TWO-SIDED], [K = 2, 3, ..., 10], [N = 3, 4, ..., 10], [α = .05, .10]

D. K DISTRIBUTIONS [ONE-SIDED], [K = 3, 4, ..., 6], [N = 3, 4, ..., 10], [α = .05, .10]

4. CORRELATION

SPEARMAN [N = 4, 5, ..., 10], [α = .05, .10, .20]

A FRESH APPROACH TO THE INTEGRATED PROCUREMENT PROBLEM VARIABLE VARIANCE

B. H. BISSINGER

INTRODUCTION. The calculation of variability for our procurement problem variable is of the utmost importance to the Navy supply system. After all, it is pivotal in setting safety level. It appears some of our best savants have taken a crack at this and the history seems to point out that one should distinguish among the following:

Models

Mathematical Statistics

Approximations

A change in any one of these may, and apparently does, affect the variance calculation.

This new approach avoids the problems others have run into.

In the appendices are fundamental formulae, a careful statistical analysis to be heeded, and a history of those attempts to solve this problem.

THE PICTURE. First, let us look at a simple, but typical, constant situation. Suppose:

L = leadtime = 5 quarters

TAT = turn-around-time = 2 quarters

D = quarterly demand = 4 units

B = regenerations per quarter = 2 units

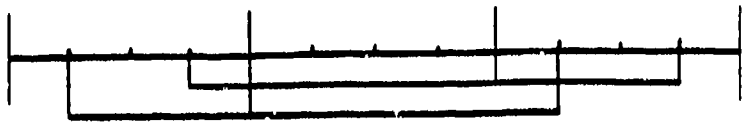
Then our net Z = procurement in a leadtime is:

$$Z = (L)(D) - (L)(B) + (B)(TAT) \text{ or}$$

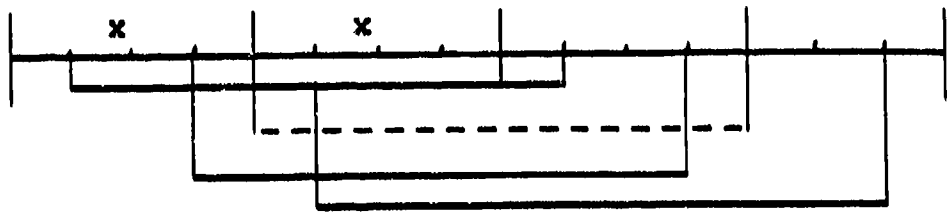
$$= D(TAT) + (D-B)(L-TAT)$$

$$= 20 - 10 + 4 = 8 + 6 = 14$$

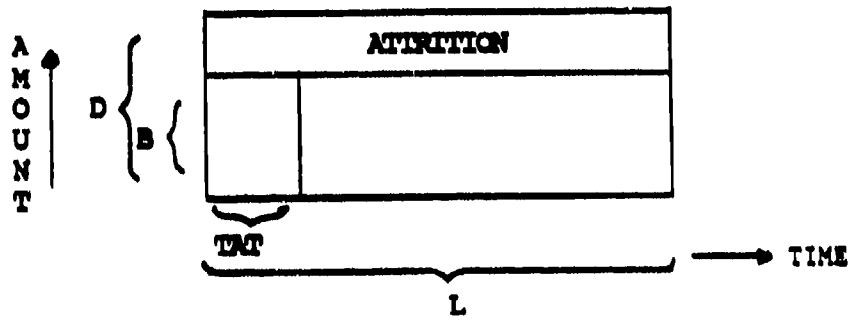
Here is a picture drawn by CDR L. Atkinson:



Let's look at a similar situation where L, TAT, and D are the same but B is increased to 3. Then Z = 20 - 15 + 6 = 8 + 3 = 11.



A similar deterministic portrayal was given by CDR T. Bunker as follows:



These mnemonic heuristic diagrams are fine if used properly to set up the relevant indeterministic expressions.

<u>THE MACHINERY</u>. Let r_1 = recovery rate and r_2 = repair rate so that r_1r_2 is the percentage (decimal equivalent) of replenishment, and hence, $1-r_1r_2$ = attrition rate.

From the just discussed and pictured process (model) we can write the

procurement problem variable as:

$$Z = Z_1 - Z_2 + Z_3$$

where

$$Z_1 = \sum_{i=1}^{L} D_i$$

$$Z_2 = r_1 r_2 \sum_{i=1}^{L} D_i$$

$$Z_3 = r_1 r_2 \sum_{i=1}^{T} D_i$$

The variance of Z is

$$\sigma_Z^2 = \sigma_{Z_1}^2 + \sigma_{Z_2}^2 + \sigma_{Z_3}^2 - 2\,\mathrm{COV}\,(Z_1, Z_2) + 2\,\mathrm{COV}\,(Z_1, Z_3) - 2\,\mathrm{COV}\,(Z_2, Z_3)$$

First let us compute the three variances:

$$E(Z_1|L) = L\mu_D \qquad V(E(Z_1|L)) = \mu_D^2\sigma_L^2$$

$$V(Z_1|L) = L\sigma_D^2 \qquad E(V(Z_1|L)) = \mu_L\sigma_D^2$$

$$\therefore \quad \mathrm{VAR}\ Z_1 = \mu_D^2\sigma_L^2 + \mu_L\sigma_D^2$$

Obviously, since Z_2 = a constant times Z_1,

$$\mathrm{VAR}\ Z_2 = r_1^2 r_2^2\ (\mu_D^2\,\sigma_L^2 + \mu_L\,\sigma_D^2).$$

Also, since Z_3 is the same as Z_1 except for T replacing L, and has the same constant multiplier as Z_2,

$$\mathrm{VAR}\ Z_3 = r_1^2 r_2^2\ (\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2)$$

Next the 2 COV (Z_1, Z_2) = 2 COV $\left[\sum_{i=1}^{L} D_i,\ r_1 r_2 \sum_{i=1}^{L} D_i \right]$

$$= 2\, r_1 r_2\ \text{VAR} \left[\sum_{i=1}^{L} D_i \right]$$

$$= 2\, r_1 r_2\ [\mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2]$$

The 2 COV (Z_1, Z_3) = 2 COV $\left[\sum_{i=1}^{L} D_i,\ r_1 r_2 \sum_{i=1}^{T} D_i \right]$

$$= 2\, r_1 r_2 \left[E\left(\sum_{1}^{L} D_i \right)\left(\sum_{1}^{T} D_i \right) - \mu_D^2 \mu_L \mu_T \right]$$

Now $E \left[\left(\sum_{1}^{L} D_i \right)\left(\sum_{1}^{T} D_i \right) | L, T \right] = LT \left[E\,(D_i^2) \right] = LT\ (\sigma_D^2 + \mu_D^2)$

Assuming L and T are independent, we get the above to be:

$$\mu_L \mu_T\ (\sigma_D^2 + \mu_D^2)$$

So the 2 COV (Z_1, Z_3) becomes:

$$2\, r_1 r_2\ [\mu_L \mu_T (\sigma_D^2 + \mu_D^2) - \mu_D^2 \mu_L \mu_T]$$

$$= 2\, r_1 r_2\ \mu_L \mu_T \sigma_D^2$$

The third covariance term follows easily from the above and we have:

$$2\ \mathrm{COV}\ (Z_2, Z_3) = 2\ r_1^2 r_2^2\ \mu_L \mu_T \sigma_D^2$$

So, combining all six terms we get:

$$\mathrm{VAR}\ Z = \mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2 + r_1^2 r_2^2\ [\mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2]$$

$$+ r_1^2 r_2^2\ [\mu_T \sigma_D^2 + \mu_D^2 \sigma_T^2]$$

$$- 2\ r_1 r_2\ [\mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2]$$

$$+ 2\ r_1 r_2\ [\mu_L \mu_T (\sigma_D^2 + \mu_D^2) - \mu_D^2 \mu_L \mu_T]$$

$$- 2\ r_1 r_2\ [\mu_L \mu_T (\sigma_D^2 + \mu_D^2) - \mu_D^2 \mu_L \mu_T]$$

$$- (1 - r_1 r_2)^2\ [\mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2]$$

$$+ r_1^2 r_2^2\ [\mu_T \sigma_D^2 + \mu_D^2 \sigma_T^2]$$

$$+ 2\ r_1 r_2\ [\mu_L \mu_T \sigma_D^2]$$

$$- 2\ r_1^2 r_2^2\ [\mu_L \mu_T \sigma_D^2]$$

$$- (1 - r_1 r_2)^2 \; [\mu_L \sigma_D^2 + \mu_D^2 \sigma_L^2]$$

$$+ r_1^2 r_2^2 \; [\mu_T \sigma_D^2 + \mu_D^2 \sigma_T^2]$$

$$+ 2 \, r_1 r_2 \, (1 - r_1 r_2) \; [\mu_L \mu_T \sigma_D^2]$$

This last formula is a model builder's dream. It has highly desirable properties. First, note the coefficients add up to unity,

$$(1 - r_1 r_2)^2 + 2 \, r_1 r_2 \, (1 - r_1 r_2) + r_1^2 r_2^2$$

$$= [(1 - r_1 r_2) + (r_1 r_2)]^2 = 1$$

So they may be considered weights attaching importance to the factors they multiply. Next, numerical values for the various factors are easily available and anyone can easily calculate the total expression.

Then it has sort of a group symmetry in that it is invariant under the transformation sending L to T, T to L and $r_1 r_1$ to $1 - r_1 r_2$ and vice versa. Molecular chemists and physicists go into ecstasy over such formulas as they say it shows strength.

Each term has meaningful sense as you read it. There is a fraction of the variance of leadtime demand, a fraction of the variance of turn-around-time demand, and an interaction term to make up the rest.

Let's say $r_1 r_2 = .9$ which I am told is not unrealistic. We get back into service 90% of what we bought after repairing. Then $1 - r_1 r_2 = .1$ and our coefficients become:

.01 on variance of leadtime demand

.81 on variance of turn-around-time demand

.18 on the interaction of the above two

It makes sense to put most of your weight on that which is most active. The interaction term can be written as

$$2\,[r_1 r_2(\mu_T \sigma_D) \cdot (1 - r_1 r_2)(\mu_L \sigma_D)]$$

which is like an association index.

FINALE. Process should always come first, like in Management Science policy should precede procedure. I owe much thanks to J. Boyarski who, after suffering with the historical presentations as I went through them, impressed me with the Markov closed loop process we have here and stressed the systems engineering aspects. I finally gave up on fiddling with what everybody else had done and started from scratch. It looks like it paid off.

Finally we see this is a true generalization of the consumable model in that if $r_1 r_2 = 0$; i.e., no repairables, we find the correct expression for a consumable.

APPENDIX A

Certain random variable expressions arise in the computations for the variance of the procurement problem variable regardless of the model. Here we give them and their variances.

1. In assuming the quarterly demands are i.i.d we compute the variance of the random variable sum of them as

$$\text{VAR} \sum_{i=1}^{L} D_i = \mu_D^2 \sigma_L^2 + \mu_L \sigma_D^2$$

Otherwise, we would have more complications. For example, if we assumed that successive demands had correlation ρ, then an additional term of the form

$$\mu_L(\mu_L - 1)\rho\ \sigma_D^2$$

would appear, thereby increasing the variance. We know the variance of a mean of correlated variables cannot be driven down by increasing sample size.

As it is, we are assuming L and D are independent.

2. For any two random variables x and y:

$$E\ (x \cdot y) = \mu_{xy} = \mu_x \mu_y + \rho_{xy}\ \sigma_x \sigma_y$$

If x and y are independent, this reduces to

$$E\ (x \cdot y) = \mu_{xy} = \mu_x \mu_y$$

3. For any two random variables x and y:

$$\sigma_{xy}^2 = E\ (xy - E\ (x \cdot y))^2$$

$$= \mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2 + \sigma_x^2 \sigma_y^2 \qquad (1)$$

$$- \rho_{xy}^2 - 2\ \mu_x \mu_y \rho_{xy} \qquad (2)$$

$$+ \text{COV}\ (x^2, y^2) \qquad (3)$$

For jointly normal with zero means COV $(x^2, y^2) = 2\,[E(xy)]^2$.

If x and y are independent this reduces to

$$\sigma_{xy}^2 = \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2 \tag{4}$$

4. COV (kx,x) = k (VAR(x)) where k is a constant.

5. COV (x, a-x) = $-\sigma_x^2$.

6. In the UICP formulation we assume the number of units demanded each time period (i) is a random variable D_i which is described by a fixed, known frequency distribution and which is not autocorrelated. Also, it's assumed the return-from-repair each time period (i) is a random variable R_i which is described by a fixed, known frequency distribution and which depends on (is correlated to) exactly one observation of demand, namely, the demand that occurred a set turn-around-time ($\bar{T}$) prior; i.e., $D_{i-\bar{T}}$. We run into the covariance of D_i and R_{i+T}. To simplify it we further assume that:

$$R_i = P_i\, D_{i-T}$$

where P_i is the return rate of the (i-T)-th period times the survival rate of the i^{th} period. Then we can write:

$$\text{COV}\,(D_i, R_{i+T}) = \text{COV}\,(D_i, P_iD_i) = E\,(D_iP_iD_i) - E\,(D_i)\,E\,(P_iD_i)$$

$$= E\,(P_i)\,E(D_i^2) - E\,(P_i)\,\,[E(D_i)]^2$$

$$= \bar{P}\,\sigma_D^2 = \frac{\bar{R}}{\bar{D}}\,\sigma_D^2$$

where we further assumed P and D are independent so that

$$E(R) = E(P)\,E(D)$$

7. NASA uses the following approximations for σ^2_{xy}:

$$\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + 2\,\rho\mu_x\mu_y\sigma_x\sigma_y$$

or

$$(\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2)(1 + \rho^2)$$

APPENDIX B

There is an essential point to be made regardless of the model used. We will illustrate it by considering three different expressions which are algebraically equivalent in deterministic algebra and also which have the same first moments when we consider the symbols to be random variables and switch to the algebra of indeterminism. However, the second moments are not necessarily equal and, hence, neither are the variances calculated therefrom.

First consider the elementary algebra identity

$$X - X = 0 \tag{1}$$

Now consider the related-in-form random variable expression

$$X_1 - X_2 \tag{2}$$

where X_1 and X_2 are i.i.d. The mean of this random variable expression is 0, and so it appears there is no need to distinguish between (1) and (2). But the variance of (2) is $2\sigma_X^2$ while the variance of a constant like 0 is 0.

Another simple example comes from taking $X + X = 2X$ and then making the variables random variables which leads to the contradiction $2\sigma_X^2 = 4\sigma_X^2$.

Why all this very elementary talk? Well, consider:

$$\sum_{i=1}^{7} D_i - \sum_{i=1}^{4} D_i \tag{3}$$

Then the variance of this is

$$\mathrm{VAR}\left(\sum_{i=1}^{7} D_i - \sum_{i=1}^{4} D_i\right) = \mathrm{VAR}\left(\sum_{i=1}^{7} D_i\right) + \mathrm{VAR}\left(\sum_{i=1}^{4} D_i\right)$$

$$= 7\sigma_D^2 + 4\sigma_D^2 = 11\sigma_D^2 \tag{4}$$

while if we use deterministic algebra first, viz,

$$\sum_{i=1}^{7} D_i - \sum_{i=1}^{4} D_i = \sum_{i=5}^{7} D_i$$

we get

$$\text{VAR}\left(\sum_{i=5}^{7} D_i\right) = 3\sigma_D^2 \tag{5}$$

So let us now consider three different expressions that exist in different presentations of our procurement problem variable. These three expressions are algebraically equivalent in deterministic algebra. Here they are:

$$\underset{(a)}{\sum_{i=1}^{L} D_i} - \underset{(b)}{\sum_{i=1}^{T} D_i} = \underset{(c)}{\sum_{i=T+1}^{L} D_i}$$

It is easily seen that if we suddenly make D_i, L and T random variables and any two D_i, D_j are i.i.d and L and D_i are independent with $\bar{L} > \bar{T}$, then the mean of (a), (b) and (c) is

$$(\bar{L} - \bar{T})\bar{D}$$

But the variances differ! Let us develop the variance of (a).

$$\text{Let } Y_a = \sum_{i=1}^{L} D_i - \sum_{i=1}^{T} D_i$$

$$E(Y_a|L,T) = L\mu_D - T\mu_D = (L - T)\mu_D$$

$$\therefore \text{VAR}\left\{E(Y_a|L,T)\right\} = \mu_D^2 \cdot \sigma^2_{(L-T)} \tag{6}$$

$$\text{Now} \qquad \text{VAR}\,(Y_a|L,T) = L\sigma_D^2 + T\sigma_D^2 \tag{7}$$

assuming the D_is are i.i.d.

Then $E(\text{VAR}\ (Y_a|L,T) = \bar{L}\sigma_D^2 + \bar{T}\sigma_D^2$ (8)

∴ The variance of Y is the sum of (6) and (8),

$$\text{VAR}\ (Y_a) = \mu_D^2\sigma^2_{(L-T)} + (\mu_L+\mu_T)\ \sigma_D^2$$

If we further assume L and T are independent, then

$$\text{VAR}\ (Y_a) = \mu_D^2\ (\sigma_L^2+\sigma_T^2) + (\mu_L+\mu_T)\ \sigma_D^2 \tag{9}$$

How about (b). Let $Y_b = \sum_{i=1}^{L-T} D_{i+T}$

$$\text{VAR}\ Y_b = (\bar{L}-\bar{T})\ \sigma^2_{D_{i+T}} + \bar{D}^2_{i+T}\ \sigma^2_{L-T}$$

$$= (\bar{L}-\bar{T})\ \sigma^2_{D_{i+T}} + \bar{D}^2_{i+T}\ (\ ^{?}_{\omega} + \sigma_T^2) \tag{10}$$

Finally the variance of (c), and we call Y_c = (c), is

$$\text{VAR}\ Y_c = \bar{L}\sigma_D^2 + \bar{D}^2\sigma_L^2 + (\bar{T} + 1)\sigma_D^2 + \bar{D}^2\sigma_T^2 \tag{11}$$
$$(?)$$

The reader will notice several similarities and dissimilarities. Before that, I call attention to the question mark under the plus sign in (11). Some places I have found a minus sign here! The variances for (a) and (c) are similar, the difference being minor and depending on integer versus continuity for T. On the other hand, the variance of (b) not only has a factor $(\bar{L}-\bar{T})$ on one term as opposed to $(\bar{L}+\bar{T})$ in (a) and (c), but it also has an in-

volved variance term which, heretofore, has been mysteriously handled. I refer to $\sigma^2_{D_T}$.

The point is that (a) and its variance are the correct approach.

APPENDIX C

Back as far as 1963 when the PARS were being written (PAR I - Application D, Operation 6 (Levels Computations for Repairables)) we find the formula for the variance of attrition demand given to be

$$\sigma^2_{D-rB} = \sigma^2_D + \sigma^2_{rB} - 2\ COV\ (rB,D) \tag{1}$$

where

D = quarterly demand

r = average repair survival rate

B = carcass return rate

σ^2_{rB} is broken down into the correct three terms, based on independence of r and B, namely,

$$\bar{r}^2\ \sigma^2_B + \bar{B}^2\ \sigma^2_r + \sigma^2_r\sigma^2_B \tag{2}$$

(See APPENDIX A - formula (4)).

Further, assuming (a) that r is independent of B and D and (b) that the RFI regenerations for a given quarter are a function of demand from a prior quarter, the expression (1) reduces to

$$\sigma^2_D + \sigma^2_{rB} - 2\ \bar{r}\ COV\ (B,D)$$

and quickly is added

$$COV\ (B,D) = \frac{\bar{B}}{D}\ \sigma^2_D .$$

Also, under the assumptions:

(a) Demand during turn-around-time is independent of attrition during leadtime less turn-around-time;

(b) Leadtime and turn-around-time are independent,

the covariance of demand during procurement turn-around-time with attrition during leadtime less turn-around-time is given to be

$$\cdot \quad \text{COV}\ (DT,(D-rB)(L-T) = -\sigma_T^2\ [\bar{D}(\bar{D} - \bar{r}\ \bar{B})]$$

Finally, the variance (V_i) of the procurement problem variable is given to be

$$V_i = \sigma^2_{DT} + \sigma^2_{(D-rB)(L-T)} + 2\ \text{COV}\ [DT,\ (D-rB)(L-T)]$$

APPENDIX D

It was in the mid-60s when we were writing the PARS that Peter Zehna turned his attention to accounting for attrition during turn-around-time.

He agreed with the others that we estimate from past history recovery rate and repair rate, say r_1 and r_2 and then $R = 1 - r_1 r_2$ is the attrition rate. Also, we all assumed that $L > T$ and that demands are mutually independent.

Initially we said the procurement problem variable Z is to account for all of the demands during a leadtime less the regenerations during that time. It was computed by accounting for the demands during turn-around-time T and adding the attritions during leadtime less turn-around-time. Zehna objected on the grounds that this implicitly assumed that regenerations for a given leadtime are a function of demands during the leadtime less turn-around-time.

He proposed what he said was more realistic and computationally simpler. He suggested we assume that regenerations are a function of the demands that occur during turn-around-time T. These regenerations are available for issue during the leadtime L and occur at a rate $r_1 r_2$. Hence, they can be expressed as the random variable

$$r_1 r_2 \sum_{i=1}^{T} D_i \tag{1}$$

So the procurement problem variable can be written

$$Z = \sum_{i=1}^{L} D_i - r_1 r_2 \sum_{i=1}^{T} D_i = R \sum_{i=1}^{T} D_i + \sum_{i=T+1}^{L} D_i \tag{2}$$

Using our usual formula for the variance of the random sum of random demands, Zehna obtained:

$$\sigma_Z^2 = R^2(\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2) + (\mu_L-\mu_T)\sigma_D^2 + \mu_D^2(\sigma_L^2 + \sigma_T^2) \tag{3}$$

$$- \sigma_D^2\,[\mu_L - (1-R^2)\mu_T] + \mu_D^2\,[\sigma_L^2 + (1+R^2)\sigma_T^2] \tag{4}$$

In our usual notation for sample estimates this gives,

$$\hat{\sigma}_Z^2 = s_D^2\,[\bar{L} - (1-R^2)\bar{T}] + \bar{D}^2\,[s_L^2 + (1+R^2)\,s_T^2] \tag{5}$$

Let's hold up here a minute and go back and rewrite (2) as:

$$Z = R\sum_{i=1}^{T} D_i + \sum_{i=1}^{L} D_i - \sum_{i=1}^{T} D_i \tag{6}$$

Then

$$\text{VAR } Z = \text{VAR}\left[R\sum_{i=1}^{T} D_i\right] + \text{VAR}\left[\sum_{i=1}^{L} D_i\right] + \text{VAR}\left[\sum_{i=1}^{T} D_i\right] \tag{7}$$

$$+ 2\,\text{COV}\left[R\sum_{i=1}^{T} D_i,\ \sum_{i=1}^{L} D_i\right] \tag{8}$$

$$- 2\,\text{COV}\left[R\sum_{i=1}^{T} D_i,\ \sum_{i=1}^{T} D_i\right] \tag{9}$$

$$- 2\,\text{COV}\left[\sum_{i=1}^{L} D_i,\ \sum_{i=1}^{T} D_i\right] \tag{10}$$

Let us compute first the three covariances.

$$\text{COV}\left[R\sum_{i=1}^{T} D_i,\ \sum_{i=1}^{L} D_i\right]$$

$$- E\left[R \sum_{i=1}^{T} D_i \times \sum_{i=1}^{L} D_i\right] - R\mu_T\mu_L\mu_D^2 \tag{8a}$$

$$\text{COV}\left[R \sum_{i=1}^{T} D_i, \sum_{i=1}^{T} D_i\right]$$

$$- R \text{ VAR} \sum_{i=1}^{T} D_i$$

$$- R\left[\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2\right] \tag{9a}$$

$$\text{COV}\left[\sum_{i=1}^{L} D_i, \sum_{i=1}^{T} D_i\right]$$

$$- E\left[\sum_{i=1}^{L} D_i \times \sum_{i=1}^{T} D_i\right] - \mu_L\mu_T\mu_D^2 \tag{10a}$$

Now let's combine (8), (9), and (10) as (8) - (9) - (10).

$$E\left[R \sum_{i=1}^{T} D_i \times \sum_{i=1}^{L} D_i\right] - R\mu_T\mu_L\mu_D^2 - R\mu_T\sigma_D^2 - R\mu_D^2\sigma_T^2 - E\left[\sum_{i=1}^{L} D_i \times \sum_{i=1}^{T} D_i\right] + \mu_L\mu_T\mu_D^2$$

$$\underbrace{- E\left[(R-1) \sum_{i=1}^{T} D_i \times \sum_{i=1}^{L} D_i\right] + (1-R)\mu_T\mu_L\mu_D^2}_{= 0} - R\mu_T\sigma_D^2 - R\mu_D^2\sigma_T^2$$

$$= - R(\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2) \quad = \quad - R \text{ VAR} \sum_{1}^{T} D_i$$

where we assumed T and L independent.

So altogether, by correct mathematical statistics, we obtain

$$\text{VAR } Z = R^2 \left[\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2\right] + \mu_L\sigma_D^2 + \mu_D^2\sigma_L^2 + \mu_T\sigma_D^2 + \mu_D^2\sigma_T^2$$

$$- 2R \left(\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2\right)$$

$$= (R^2-2R+1) \left[\mu_T\sigma_D^2 + \mu_D^2\sigma_T^2\right] + \mu_L\sigma_D^2 + \mu_D^2\sigma_L^2$$

Separating this into two terms, one on σ_D^2 and one on μ_D^2, as Zehna did, yields

$$\text{VAR } Z = \left[(R^2-2R+1)\mu_T + \mu_L\right]\sigma_D^2 + \left[\sigma_L^2 + (R^2-2R+1)\sigma_T^2\right]\mu_D^2$$

$$= \left[\mu_L + (1-R)^2\mu_T\right]\sigma_D^2 + \left[\sigma_L^2 + (1-R)^2\sigma_T^2\right]\mu_D^2$$

We note this is very similar to Zehna's result (4). The difference lies in the coefficients

$$(1-R)^2 \text{ vs } 1-R^2$$

and

$$(1-R)^2 \text{ vs } 1+R^2$$

So we see that Zehna's coefficient on $\mu_T\sigma_D^2$ is negative and, hence, makes a smaller coefficient. On the other hand, his coefficient on $\mu_D^2\sigma_T^2$ is larger by 2R.

In 1964, J. W. Prichard of BUSANDA Navy Headquarters (today NAVSUP) presented a paper entitled "Inventory Model for Repairable Items - Theory and Practices."

He let $Z_2 = DL - BL + BT - DT + (D-B)(L-T)$ be the random variable of the amount of material demanded in a leadtime by users and by the repair process, but not satisfied by the repair process. The variance of Z_2 becomes:

$$VAR\{Z_2\} = \sigma^2_{DT} + \sigma^2_{(D-B)(L-T)} + 2\ COV\ [DT,\ (D-B)(L-T)]$$

The last term, which is $-2\sigma^2_T\ [\bar{D}(\bar{D}-\bar{B})]$, is needed because of the obvious correlation between gross demand during a turn-around-time and the net demand to be met from purchase during that portion of the procurement leadtime in excess of the turn-around-time.

The other two terms in the expression for VAR $\{Z_2\}$ can be expanded into the form for sums over a random interval of random demands, viz

$$\sigma^2_{DT} = \bar{T}\sigma^2_D + (\bar{D})^2\sigma^2_T$$

$$\sigma^2_{(D-B)(L-T)} = (\bar{L}-\bar{T})\ [\sigma^2_D + \sigma^2_B + 2\ COV\ (D,B)]$$

$$+ (\bar{D}-\bar{B})^2\ [\sigma^2_T + \sigma^2_L]$$

The covariance term 2 COV (D,B) is approximately equal to

$$-\ 2\ \frac{\bar{B}}{\bar{D}}\ \sigma^2_D$$

So we end up with

$$VAR\{Z_2\} = \bar{T}\sigma^2_D\ (\bar{D})^2\ \sigma^2_T + (\bar{L}-\bar{T})\left[\sigma^2_D + \sigma^2_B - 2\ \frac{\bar{B}}{\bar{D}}\ \sigma^2_D\right]$$

$$+ (\bar{D}-\bar{B})^2\ [\sigma^2_T + \sigma^2_L] - 2\ \bar{D}(\bar{D}-\bar{B})\sigma^2_T$$

Note that B is used here for rB in the PARS example. This same approach and results were used by J. Schnelker.

Here is a development by CDR Keith Lippert without the covariance term:

$$\text{VAR}\,[D\times T + (D-rB')(L-T)] = V(D\times T) + V(D-rB')(L-T)$$

$$= \bar{D}^2\sigma_T^2 + \bar{T}\sigma_D^2 + (D-rB')^2\sigma_{L-T}^2 + (\bar{L}-\bar{T})\sigma_{D-rB'}^2$$

$$= \bar{D}^2\sigma_T^2 + \bar{T}\sigma_D^2 + (D-rB')^2(\sigma_L^2+\sigma_T^2) + (\bar{L}-\bar{T})\sigma_{D-rB'}^2$$

$\sigma_{D-rB'}^2 = \sigma_D^2 + \sigma_{rB'}^2$, assuming independence.

$$\sigma_{rB'}^2 = \int_0^\infty \int_0^\infty (rB'-\mu_r\mu_{B'})^2 f(r)f(B')\,dr\,dB'$$

$$= \int_0^\infty \int_0^\infty (r^2B'^2 - 2rB'\mu_r\mu_{B'} + \mu_r^2\mu_{B'}^2) f(B')f(r)\,dr\,dB'$$

$$= \int_0^\infty r^2E(B')^2 f(r)\,dr - \int_0^\infty 2r\mu_r\mu_{B'}^2 f(r)\,dr + \int_0^\infty \mu_r^2\mu_{B'}^2 f(r)\,dr$$

$$= E(r^2)\,E(B'^2) - 2\mu_r^2\mu_{B'}^2 + \mu_r^2\mu_{B'}^2$$

$$= (\sigma_r^2 + \mu_r^2)(\sigma_{B'}^2 + \mu_{B'}^2) - \mu_r^2\mu_{B'}^2$$

$$= \sigma_r^2\sigma_{B'}^2 + \sigma_r^2\mu_{B'}^2 + \mu_r^2\sigma_{B'}^2$$ (this follows from our APPENDIX A equation (3).

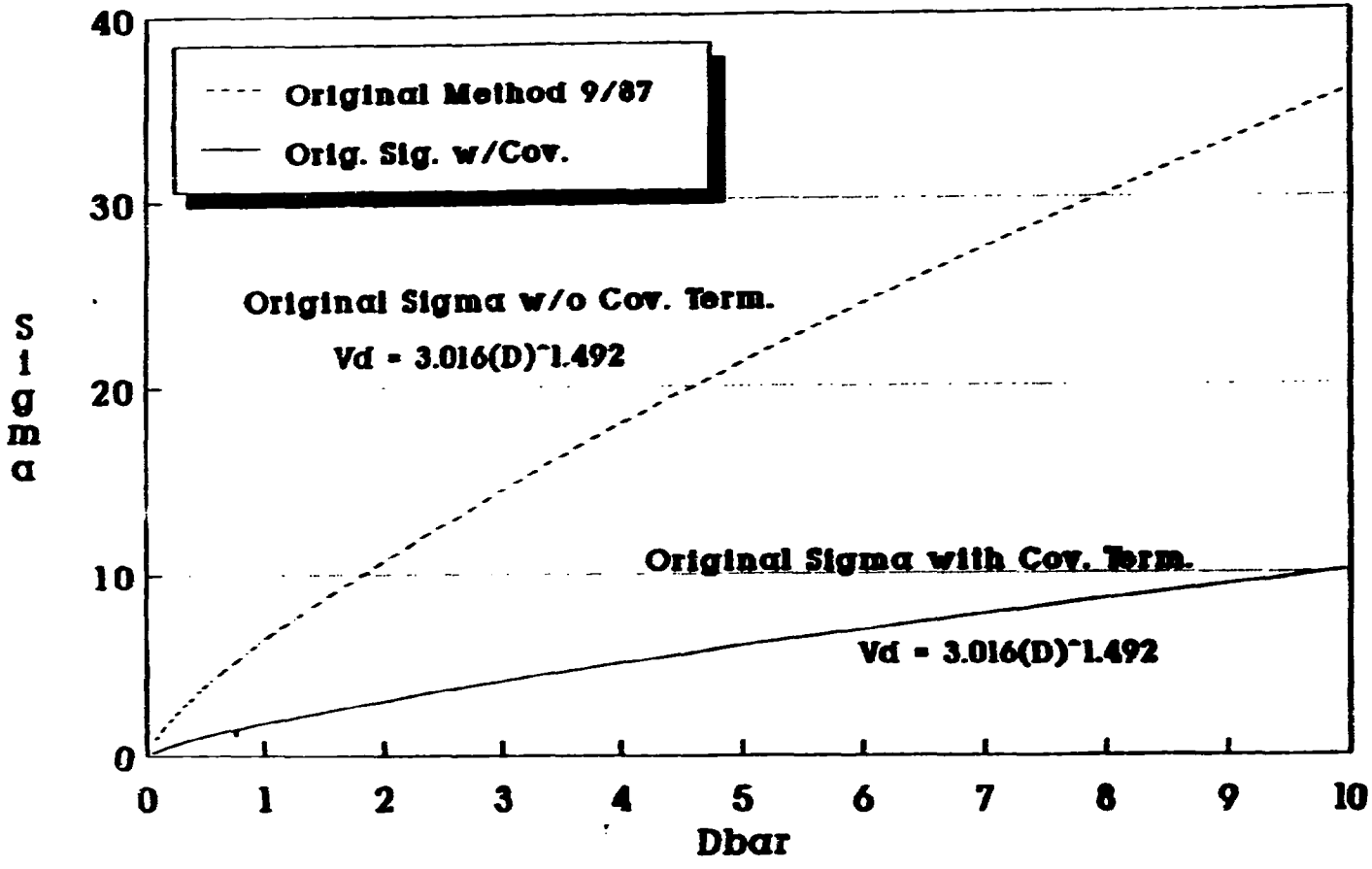
So in total for Z = procurement problem variable

$$V(Z) = \overline{D}^2\sigma_T^2 + \overline{T}\sigma_D^2 + (\overline{D}-\overline{rB'})^2(\sigma_L^2 + \sigma_T^2) + (\overline{L}-\overline{T})\ [\sigma_D^2 + \sigma_r^2\sigma_{B'}^2 + \sigma_r^2\mu_{B'}^2 + \mu_r^2\sigma_{B'}^2]$$

Compressing rB' into simply B, this becomes

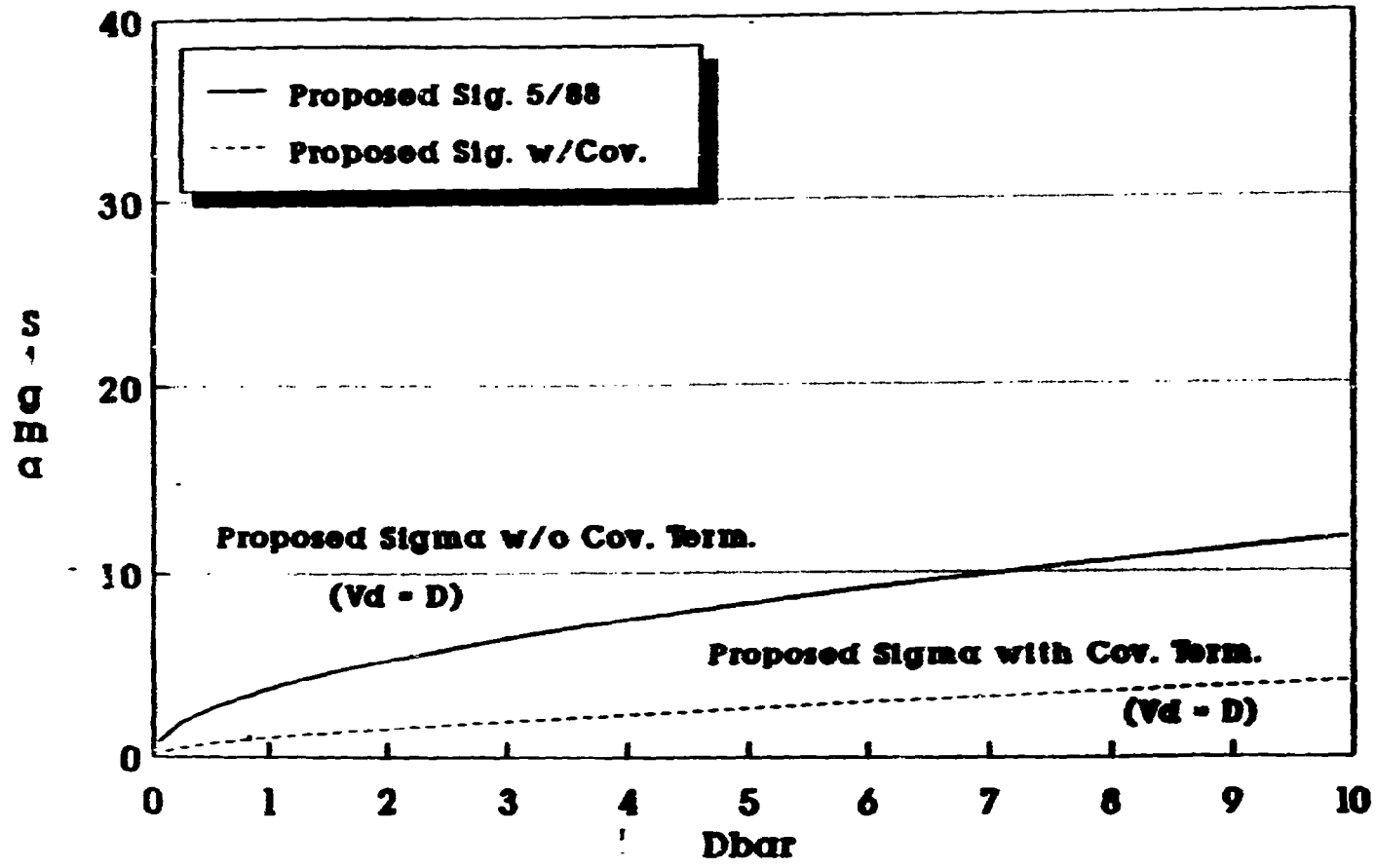
$$\overline{D}^2\sigma_T^2 + \overline{T}\sigma_D^2 + (\overline{D}-\overline{B})^2(\sigma_L^2 + \sigma_T^2) + (\overline{L}-\overline{T})(\sigma_D^2 + \sigma_B^2)$$

SIGMA VS. Dbar
(Original Sigmas)



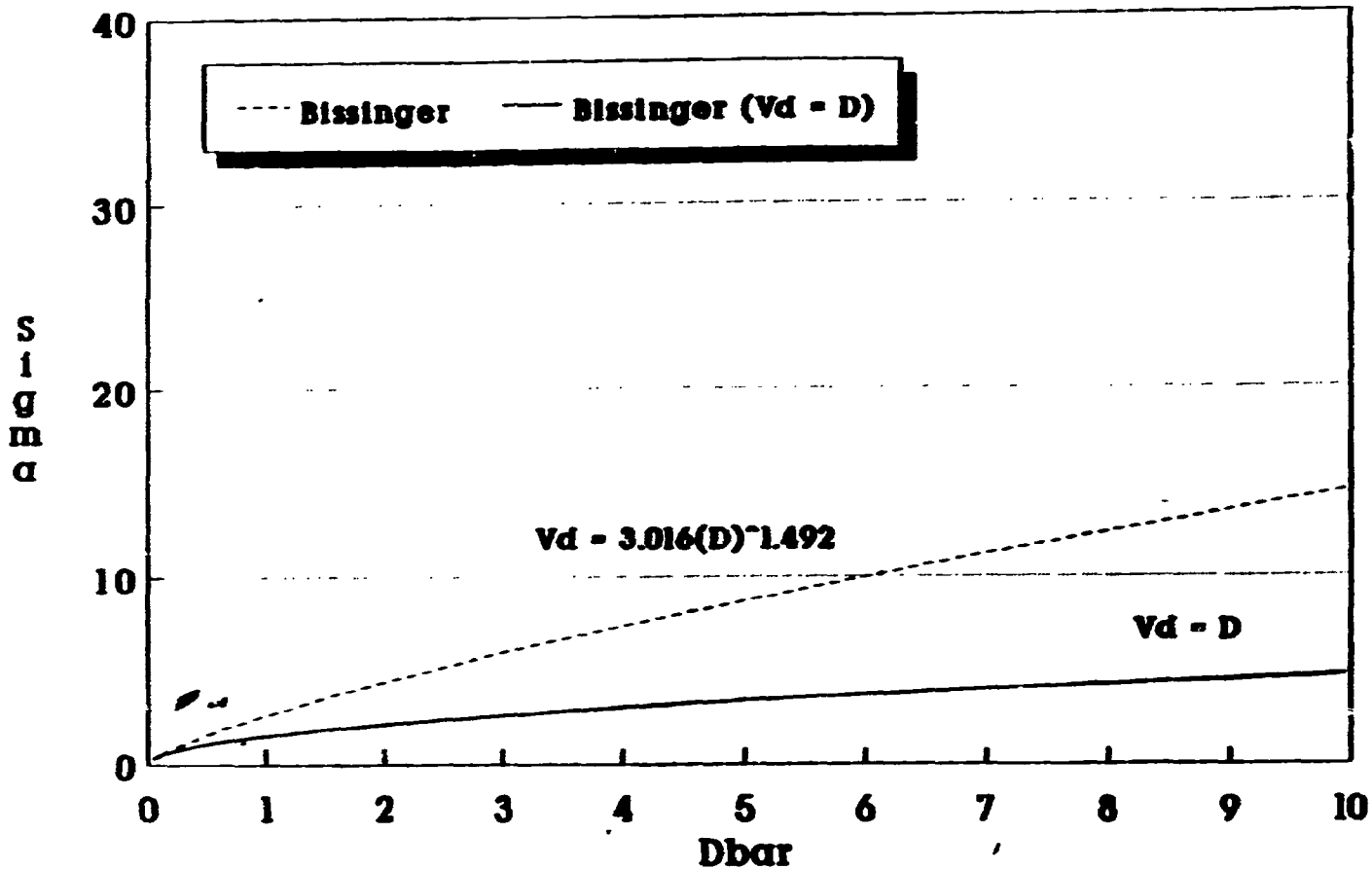
All Variables Constant Except Vd.

SIGMA VS. Dbar
(Proposed Methods)



All Variables Constant Except Vd.

SIGMA VS. Dbar
(Bissinger Method)



All Variables Constant Except Vd.

LATIN HYPERCUBE SAMPLING: A WAY OF SAVING COMPUTER RUNS

W. J. Conover
College of Business Administration
Texas Tech University
Lubbock, Texas 79424

ABSTRACT. When real-life situations are modeled using a computer program, the computer program is frequently very large and takes a long time to make each run. In order to get the most information from a limited number of computer runs, latin hypercube sampling was invented. The wide-spread usage of latin hypercube sampling attests to its value in producing precise estimates of the output distribution parameters. In addition, a useful method for inducing correlations among the input variables in simulations is discussed.

1. INTRODUCTION. The advent of high-speed computers has opened new doors for solving difficult real-world problems. Computer codes are written to simulate the behavior of the real-world situation, and then the codes are run repeatedly on the computer to estimate the outcome under various different circumstances, where those circumstances are used as inputs to the computer code. Unfortunately, these computer codes often become very complex in an attempt to make the codes as realistic as possible, and as a result they take so long to run on the computer that the number of runs is limited by time and money constraints. Also, computer codes become more complex when the number of different input variables increases.

Thus the following situation often arises. A complex computer code is written that mimics the real life situation as well as one can expect from any computer code. It contains many, perhaps hundreds, input variables or parameters that can be varied to represent different circumstances that should be considered, and it takes so long to run on the computer that only a few simulation runs (say 20 to 100) are possible due to time and money constraints.

How is this possible? In everyone's mind there's the feeling that the number of runs must be larger than the number of variables. However, that notion comes from solving systems of linear equations, and does not apply to computer runs. For example, one could simply choose a likely value for each of the k input variables, and make a single computer run using these values. Then one could use a different set of values for the input variables, perhaps representing a possible undesirable scenario, and make a second run on the computer. So k, the number of input variables, can be much larger than n, the number of runs.

The question then becomes, how should the various values of the input variables be selected so as to get the most information, in some sense, out of a limited number of runs? One approach is the deterministic approach, which says to select particular sets of values of the input variables that you, or someone else, want to examine for one reason or another. The output of the computer code then applies to the scenarios represented by those sets

of input values. There are obvious advantages to this approach, but the main disadvantage is that few, if any, probability statements can be made, and often any kind of post-hoc analysis is very limited.

A second approach is to use a monte carlo approach, and randomly select values for each input variable, one value at a time, and do the same for all input variables. This assumes that each input variable has a known probability distribution so a random selection may be made. Then the output is one random value of the output. By repeating the procedure several times, several independent random observations are made on the output, and estimates of the output probability distribution can be made. This method is called random sampling. It allows for many different types of probability statements on the output, or concerning the relative importance of the various input variables.

A third approach, called latin hypercube sampling, is discussed in this paper. It has been used for at least ten years by several national research laboratories, notably Los Alamos National Laboratories and Sandia Laboratories. It is used in at least 22 different countries for selecting input variables in long-running computer codes, primarily for modeling nuclear reactor behavior, and the behavior of deep underground nuclear waste repositories. Inquiries regarding a computer code that facilitates its usage should be addressed to Dr. Ronald L. Iman, Sandia Laboratories, Albuquerque, (505)844-8834, who has gone out of his way in the past to make this program available to prospective users.

The popularity of latin hypercube sampling is due to its characteristic of having a relatively small variance, as compared with random sampling for example, in the estimates of the output distribution. Thus the same types of probability statements available from random sampling are also available using latin hypercube sampling, but usually with much more precision.

2. LATIN HYPERCUBE SAMPLING. One characteristic of most computer-coded models with many input variables is that some input variables are more influential than others in affecting the outcome. We would concentrate our attention on the more influential input variables, if only we knew which ones they were. But that is often the purpose of the simulation, to find out which input variables are the most influential on the outcome.

If we knew that the outcome was almost entirely dependent on one input variable, say X_1, then we would almost certainly want to select values of X_1 that span its entire range. In this way we could see how the outcome varies over the entire range of values of X_1, and we would have a complete picture of the model's behavior. If we were allowed to make n runs on the computer, we could divide the range of X_1 into n intervals of equal length and select one value from each interval for each run. Some of the intervals may be very unlikely to experience in real life, however, and besides that, what do we do if the range of X_1 is infinite? So it makes more sense to divide the range of X_1 into n intervals of equal probability, rather than of equal length, and randomly sample one value from each interval. Thus all of the n values of X_1 carry the same weight, and no problem arises if the range of X_1 is infinite.

The problem is that we don't know, before running the code, which variable is the most important. Furthermore, in many situations there is more than one output from the model, and while X_1 may be the most important input variable for output Y_1, say, another input variable X_2 may be the most influential input variable for another output Y_2, say. Or if the output is a function of time, one input variable may be the most influential one at an early point in time, while another one may be the most influential one at a later point in time. In fact this is the rule more than the exception. How do we handle this situation?

One obvious solution is to treat both X_1 and X_2 with equal consideration. Stratify over the entire range of X_1 to obtain the n values of X_1 as described above, and in a similar manner stratify over the entire range of X_2 to obtain the n values of X_2 for the n computer runs. Then how do we decide which values of X_1 to pair with the values of X_2 in the various computer runs? The approach used in this section is simply to pair them in a random manner, as variables would be paired in real life if they were independent of each other. In the next section a method of pairing is discussed, to achieve a desired correlation between X_1 and X_2. But for now, random pairing is used.

Of course it now becomes obvious what to do if a third input variable X_3 is also important. Stratify over the entire range of X_3 to get the n input values for X_3, and do a random permutation of those n values to match them with the (X_1, X_2) pairs already established. A similar treatment can be made of all of the input variables. In that way if one of them turns out to be very important, it has been treated with importance by stratifying over its entire range. If it turns out that one of the input variables is of little or no importance in influencing the output, nothing is lost using this procedure since all of the influential input variables are stratified over their entire range. Including this unimportant variable neither aids nor inhibits the amount of information obtained from the other variables.

Intuitively this seems like an efficient method for getting the most information out of a limited number of computer runs, but how good is it really? In an attempt to answer this question several different sampling plans were compared using real computer codes, by McKay, Conover and Beckman (1979), Iman, Conover and Campbell (1980) and Iman and Conover (1980). In all cases the output parameters were estimated with much more precision using latin hypercube sampling than with any of the other procedures examined, and the improvement was dramatic. This does not imply that there are not better methods for selecting input variable, or that this same dramatic improvement will be evident for all types of computer codes. It was true for the codes we examined, when compared with random sampling and a different form of stratified sampling.

One disadvantage of latin hypercube sampling is that even though the estimates are very precise, no measure of the precision is available, as it is when using random sampling. The solution to this problem lies in replicating a latin hypercube sample several times. For example, if a total of 100 runs is allowed on the computer, first use 10 runs, or 20 runs if you prefer, for a latin hypercube sample, where each variable is stratified over

10 (or 20) intervals. Then repeat the procedure for another 10 runs, again stratifying over 10 intervals for each variable, but of course the individual values are unlikely to be the same as before, and the random matching of one variable with another is unlikely to be the same as before. By repeating this procedure until the total number of runs is exhausted, several independent estimates of the output are obtained, where each estimate has the precision one can expect from latin hypercube sampling, and the group of estimates together provide an estimate of that precision. This variation of latin hypercube sampling is explored by Iman and Conover (1980), and as one would expect some precision is lost by this combination of latin hypercube sampling and random sampling, but the benefit is in obtaining a measure of the precision in the form of a standard deviation of the estimate. The new level of precision is somewhere between pure latin hypercube sampling and pure random sampling.

3. CORRELATING THE INPUT VARIABLES. Thus far it has been tacitly assumed that the input variables are mutually independent, and therefore the population correlation matrix is the identity matrix I.

$$I = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

The sample correlation matrix, the matrix of sample correlation coefficients representing the actual correlation of the selected input values for the various input variables, will be close to I, with differences due solely to sampling variability.

Often the input variables in a computer code represent variables which in real life are correlated. If the input variables in the computer code had a sample correlation close to the real correlation between those variables, the result would be a more realistic simulation, with more believable results. How can we match the input variables so that the matching is no longer random, but rather contrived to achieve a target correlation? The method described in this section shows how to achieve a target rank correlation, which may be the closest we can come to achieving a target correlation due to the possibility of long-tailed input distributions where outlying observations dominate the regular correlation coefficient, but have minimal effect on the rank correlation coefficient. Recall, the rank correlation coefficient, called Spearman's correlation coefficient, is just the regular correlation coefficient computed on the ranks of the observations. See Conover (1980) for a complete description of rank correlation.

An example can help describe the concept. Suppose n = 15 runs are authorized on a model with k = 6 input variables. Three of the input variables are mutually independent, and the other three are highly correlated. The population correlation matrix C looks like this.

10 (or 20) intervals. Then repeat the procedure for another 10 runs, again stratifying over 10 intervals for each variable, but of course the individual values are unlikely to be the same as before, and the random matching of one variable with another is unlikely to be the same as before. By repeating this procedure until the total number of runs is exhausted, several independent estimates of the output are obtained, where each estimate has the precision one can expect from latin hypercube sampling, and the group of estimates together provide an estimate of that precision. This variation of latin hypercube sampling is explored by Iman and Conover (1980), and as one would expect some precision is lost by this combination of latin hypercube sampling and random sampling, but the benefit is in obtaining a measure of the precision in the form of a standard deviation of the estimate. The new level of precision is somewhere between pure latin hypercube sampling and pure random sampling.

3. CORRELATING THE INPUT VARIABLES. Thus far it has been tacitly assumed that the input variables are mutually independent, and therefore the population correlation matrix is the identity matrix I.

$$I = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

The sample correlation matrix, the matrix of sample correlation coefficients representing the actual correlation of the selected input values for the various input variables, will be close to I, with differences due solely to sampling variability.

Often the input variables in a computer code represent variables which in real life are correlated. If the input variables in the computer code had a sample correlation close to the real correlation between those variables, the result would be a more realistic simulation, with more believable results. How can we match the input variables so that the matching is no longer random, but rather contrived to achieve a target correlation? The method described in this section shows how to achieve a target rank correlation, which may be the closest we can come to achieving a target correlation due to the possibility of long-tailed input distributions where outlying observations dominate the regular correlation coefficient, but have minimal effect on the rank correlation coefficient. Recall, the rank correlation coefficient, called Spearman's correlation coefficient, is just the regular correlation coefficient computed on the ranks of the observations. See Conover (1980) for a complete description of rank correlation.

An example can help describe the concept. Suppose n = 15 runs are authorized on a model with k = 6 input variables. Three of the input variables are mutually independent, and the other three are highly correlated. The population correlation matrix C looks like this.

$$C = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & .75 & -.70 \\ 0 & 0 & 0 & .75 & 1 & -.95 \\ 0 & 0 & 0 & -.70 & -.95 & 1 \end{bmatrix} \end{array}$$

Each input variable has 15 values, obtained by using the stratification procedure described for latin hypercube samples. If the 15 values for each input variable are permuted randomly the sample correlation matrix might look like this.

$$T = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} 1.00 & .10 & -.47 & -.23 & .26 & .17 \\ .10 & 1.00 & -.31 & .07 & .48 & -.23 \\ -.47 & -.31 & 1.00 & .34 & -.20 & .19 \\ -.23 & .07 & .34 & 1.00 & -.04 & -.03 \\ .26 & .48 & -.20 & -.04 & 1.00 & .05 \\ .17 & -.23 & .19 & -.03 & .05 & 1.00 \end{bmatrix} \end{array}$$

The matrix T shows how random correlations may differ from the target value of zero, and sometimes the difference is fairly large. In this case the target correlations are given in the matrix C. How can one obtain correlations, albeit rank correlations, close to the ones in C?

If the values of the input variables are permuted so that their <u>rankings</u> agree with the following rankings, then their rank correlation coefficients will be given by the rank correlation matrix M, given below.

Run Number		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	X_1	15	3	5	13	14	9	2	8	10	6	1	7	11	12	4
Ranks of	X_2	15	6	12	8	5	1	4	3	7	9	13	2	11	10	14
Variables	X_3	5	10	4	7	14	1	2	8	13	6	12	15	3	11	9
	X_4	2	1	15	10	11	8	9	6	12	3	13	5	14	7	4
	X_5	1	3	15	12	13	10	5	6	14	8	9	2	7	11	4
	X_6	15	13	1	2	7	8	11	12	3	6	4	10	9	5	14

$$M = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} 1.00 & .02 & .05 & .04 & .22 & -.08 \\ .02 & 1.00 & -.06 & .08 & -.01 & -.05 \\ .05 & -.06 & 1.00 & -.08 & .05 & -.11 \\ .04 & .08 & -.08 & 1.00 & .73 & -.70 \\ .22 & -.01 & .05 & .73 & 1.00 & -.89 \\ -.08 & -.05 & -.11 & -.70 & -.89 & 1.00 \end{bmatrix} \end{array}$$

Note how close the rank correlations are to the target correlations given above in the matrix C. Even the correlations aiming at the value zero come much closer to zero than the random correlations in the matrix T. Thus even if the input variables are independent, one may prefer to use this procedure to obtain nearly orthogonal (in the sense of ranks) input vectors, rather than relying on random matching which may produce, by chance, correlations quite far from the target values of zero, as shown in the matrix T.

It is necessary for the number of runs n to be larger than the number of variables k* for which correlations are being designated, in order to use this procedure. Note that k* may be less than the total number of variables k.

One advantage of using the rank correlation coefficients becomes apparent. The ranks, when paired as they are above, always result in the rank correlation matrix M, no matter what the original numbers are, and therefore no matter what the marginal distributions might be. Thus this method of inducing rank correlations is free of any distributional assumptions regarding the input variables.

Although we are using this method of inducing correlations in conjunction with latin hypercube samples, it is in no way tied to latin hypercube sampling. It works equally well with random sampling, or any other way of obtaining values for the input variables. All that is required is a rearrangement of the input values so that their ranks agree with a prescribed set of ranks, in order to obtain a rank correlation matrix close to the target rank correlation matrix.

Of course the big question is, how does one obtain the prescribed set of rankings for any given rank correlation matrix, as given above for the matrix M? As you would expect, the method is not simple. It can be done by hand, but the Sandia computer program is recommended for convenience, since it takes the difficulty out of the procedure. For those who are not afraid of matrix manipulation, the procedure is as follows.

1. Start with any set of n numbers, called scores, where n is the number of runs. We usually use normal scores, which are the $i/(n+1)$ quantiles from a standard normal distribution, $i = 1, \ldots, n$, which are readily available from any table of the standard normal distribution such as that in Conover (1980). Denote those scores by $a(1), \ldots, a(n)$.

2. Form a matrix R with k* columns in it, where each column contains a random permutation of the n scores, and where k* represents the number of input variables being correlated. Be sure all permutations are distinct.

3. Find the sample correlation matrix T of R. Note that T is the regular correlation matrix, not the rank correlation matrix. However it is a characteristic of normal scores, and normal random variables, that regular correlation coefficients and rank correlation coefficients are usually quite similar.

4. Find a matrix Q such that $QQ' = T$, where Q' denotes the transpose of Q. Mathematicians have devised several methods for finding Q. The one that we use is the Cholesky factorization scheme, which results in a lower triangular matrix for Q.

5. Let the target correlation matrix be denoted by C. Find a matrix P such that $PP' = C$. Again, we use the Cholesky factorization scheme because of its relative simplicity.

6. Find $S = PQ^{-1}$ and compute $R^* = RS'$. The ranks of the matrix R^* (one column at a time) are the ranks we are seeking. Any set of input vectors with the same ranks as R^* will have a rank correlation matrix close in value to target correlation matrix C.

Why does this work? First, the regular sample correlation matrix of R^* is C. This is a simple result that can be shown with a little matrix algebra. Second, because we started with normal scores, the rank correlation coefficients of R^* are usually numerically close to the regular correlation coefficients, given in C. Therefore any matrix with the same ranks as R^* will have the same rank correlations as R^*, which should be close to C.

REFERENCES

Conover, W.J. (1980), Practical Nonparametric Statistics, Second Edition, John Wiley and Sons, New York.

Iman, R.L., and Conover, W.J. (1980), "Small Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment," Communications in Statistics, Theory and Methods, Volume A9, No. 17, pp.1749-1874.

Iman, R.L., and Conover, W.J. (1982), "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," Communications in Statistics, Simulation and Computation, Volume 11, No. 3, pp.311-334.

Iman, R.L., Conover, W.J., and Campbell, J.E. (1980), "Risk Methodology for Geologic Disposal of Radioactive Waste: Small Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment," NUREG/CR-1397, SAND 80-0020, Sandia Laboratories.

McKay, M.D., Conover, W.J., and Beckman, R.J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, Vol. 21, No. 2, pp. 239-245.

GRAPHICAL TOOLS FOR EXPERIMENT DESIGN

Russell R. Barton
School of Operations Research and Industrial Engineering
Cornell University
Ithaca, New York 14853

ABSTRACT

Graphical methods for designing experiments have been used since the inception of statistical experiment design, yet this approach has received little recognition in the literature. This presentation surveys historical uses of graphical displays and shows how graphical representations can clarify the difference between a bad design and a good one. Some practical rules for generating new designs by graphical means are presented.

KEYWORDS: Experiment Design, Graphical Methods

I. INTRODUCTION

How can graphical tools be used in the process of designing an experiment? First, consider the steps involved in experiment design. One can think of this process as composed of five steps. These must occur before any data are collected, and before statistical analyses are performed. They are:

1, define the purpose of the experiment,
2, identify the independent, intermediate, dependent, and nuisance variables,
3, classify the variables as quantitative or qualitative, linear or nonlinear effect (independent variables), and fixed or varied during the experiment (independent variables),
4, using the above information, choose or create a design, and
5, validate the design.

This paper presents graphical methods for steps 2, 4, and 5 of this process. For step 2, we will show Andrews and fishbone diagrams. Multidimensional point plots and a variety of other techniques can be used in step 4. For step 5, we will discuss graphical properties of good designs, and the importance of checking projections.

Because of the high graphical content of this presentation, the format of the following paper is unconventional. Its form is more like that of an oral presentation, with figures placed on the left side of each page, and the accompanying text on the right (opposite each figure)[1]. This allows approximately sixty figures to be discussed in thirty pages, which might otherwise have taken twice the space.

[1] The following pages come from a session entitled "Practical Graphical Techniques for the Design and Analysis of Experiments" presented by James Filliben, Gerald Hahn, and this author at the 1987 American Statistical Association Winter Conference in Orlando, Florida. These figrues are more complete than the Army Design of Experiments presentation in most ways, although some recent material was presented in Monterey that is missing here.

VIEWGRAPH

TEXT

1

Practical Graphical Techniques for the Design and Analysis of Experiments

Russell R. Barton
James J. Filliben
Gerald J. Hahn

PART 1 : GRAPHICAL DOX

PART 2 : GRAPHICAL ANALYSIS

PART 3 : RECENT APPLICATIONS

- OVERVIEW OF A THREE PART TALK
- ABOUT 20 MINUTES PER SECTION
- BEGIN WITH GRAPHICAL METHODS FOR DESIGN--NOT JUST FOR VIEWING DESIGNS, BUT FOR DESIGNING DESIGNS.
- WHY GRAPHICAL METHODS?
  >provide better understanding of design
  >make it easy to generate a new design
  >provides a layout to run the design from
- START FROM A BROAD CONTEXT:

  WHAT ARE THE EVENTS LEADING TO THE NEED FOR AN EXPERIMENT?

2

"WE EMPHASIZE THAT THE SELECTION OF THE MATRIX OF EXPERIMENTAL POINTS REPRESENTS ONLY THE PROVERBIAL TIP OF THE ICEBERG. THEREFORE, WE STRESS SUCH MATTERS AS THE NEED FOR CLEARLY DEFINING THE GOAL OF THE TEST PROGRAM, ENUMERATING ALL POSSIBLE VARIABLES, AND HOW TO HANDLE THEM."

-G.J. Hahn

"WHAT IS THE OBJECTIVE OF THIS INVESTIGATION ?"

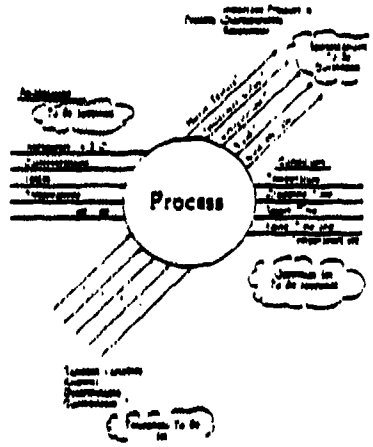
-J.S. Hunter

- Why is the experiment necessary?
- What is known about the system that is being investigated?
- What are the KEY VARIABLES:
  Independent
  Dependent
  Intermediate
- Anticipated complexity of relationships?
- Known constraints on:
  variable/factor values
  experimental procedure
- What is the expected outcome?
- Why use GRAPHICAL methods?
  right-brain, creative
  powerful, robust

VIEWGRAPH

TEXT

3



What do we mean by GRAPHICAL designs?

Andrews used representations that were graphic indeed!

They convey more than just the combinations of factor levels that will be tried, triggering the viewer's imagination to think about the often important details as well as the main structure (cf viewgraphs 53&54)

At the first level of experiment design, one needs to view the process that will be investigated. This viewgraph shows the representation Andrews used to plan experiments for a meat processing operation.

Source: Andrews (1964).

4

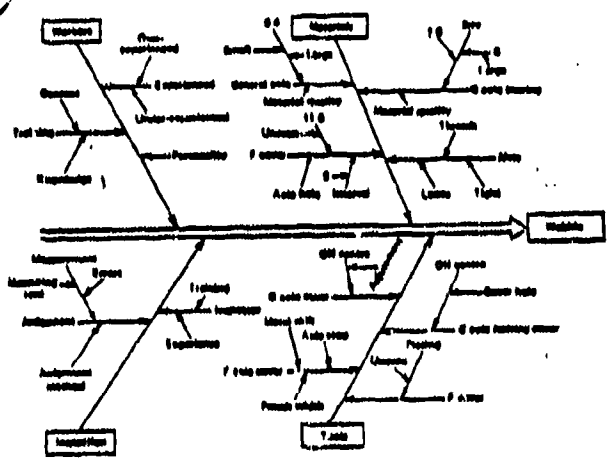


Figure 3.8 Cause-and-effect diagram for wobbling (dimension analysis)

Ishikawa's "fishbone" diagrams: quicker to draw, help to identify appropriate experiments to try.

Several forms:
- cause-effect
- process-oriented
- clustered lists

A process-oriented diagram for the axle manufacturing problem would be organized to have the major process steps on the backbone, with subprocesses hanging off these, etc. Causes of wobble would tend to be the outermost 'bones' on the 'skeleton'

Source: Ishikawa (1982).

GRAPHICAL DESIGN OF EXPERIMENTS -- R. BARTON

VIEWGRAPH

TEXT

5

Outline

What graphical tools exist to aid in designing experiments?

What graphical concepts do these tools exploit?

What are the strengths and limitations of graphical methods?

What role can computers play in graphical DOX?

Summary -- the place of graphical methods in DOX

The following pages show graphical methods to address specific kinds of designs, e.g. factorial, lifetest, etc.

Greatest concentration on multidimensional point plots for factorial and fractional designs. Reason: the ratio

$$\frac{\text{practical value}}{\text{current use}}$$

Basic outline of the DOX portion of this talk is at left.

6

13.1 A LIST OF CONSTRUCTION METHODS

The following methods of constructing factorial designs literature:

(i) Orthogonal arrays.
(ii) Balanced arrays.
(iii) Latin squares and orthogonal Latin squares.
(iv) Hadamard matrices.
(v) Finite geometries.
(vi) Confounding.
(vii) Group theory.
(viii) Algebraic decomposition.
(ix) Combinatorial topology.
(x) Foldover.
(xi) Collapsing of levels.
(xii) Composition (direct product and direct sum).
(xiii) Codes.
(xiv) Block designs.
(xv) *F*-squares.
(xvi) Weighing designs.
(xvii) Lattice designs.
(xviii) Finite graphs.
(xix) One-at-a-time.

Graphical methods for DOX not recognised as an entity historically. Computerized literature search gave ZERO titles, keywords in past 10 years with both GRAPHICAL and DOX.

Source: Raktoe, et. al. (1981).

VIEWGRAPH

TEXT

7

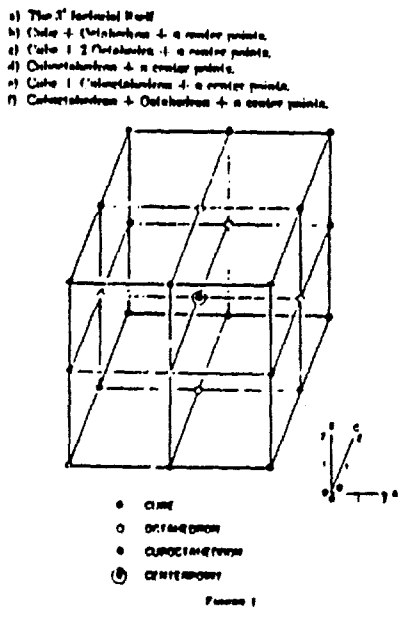
"Definition 13.1: A txn matrix A with entries from a set S of s symbols is called an orthogonal array of size n, t constraints, s levels, strength d, and index λ if any dxn submatrix of A contains all s^d possible dx1 column vectors based on s symbols of S with the same frequency λ."

-Raktoe, Heydayat, and Federer

THE MAIN POINT: it is easier to understand, manipulate and create experiment designs when they are represented graphically. Mathematical descriptions can be precise, sometimes clear, rarely easy to manipulate.

Source: Raktoe, et. al. (1981).

8



First volume, first paper in Technometrics, primary journal for examples of graphical DOX.

Several important concepts that will occur again in later viewgraphs:

1) designs decompose into subsets

2) vertices of regular polyhedra make good point subsets

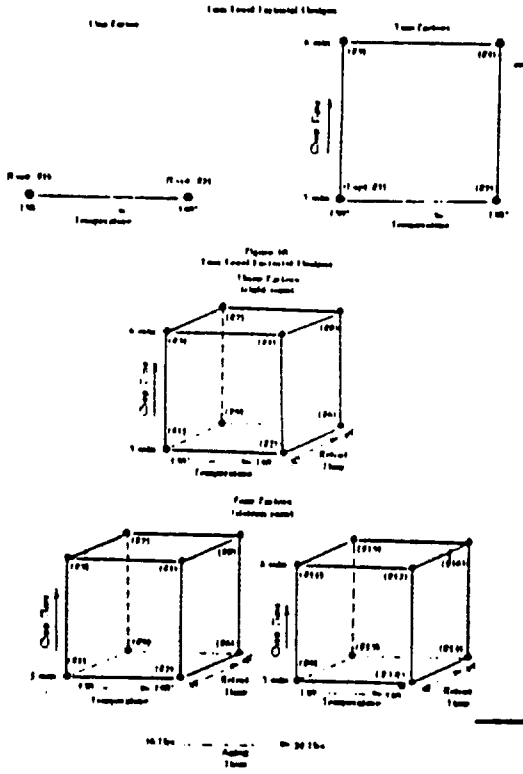
3) use of point symbols to add information to the plot

source: DeBaun (1959).

VIEWGRAPH

TEXT

9

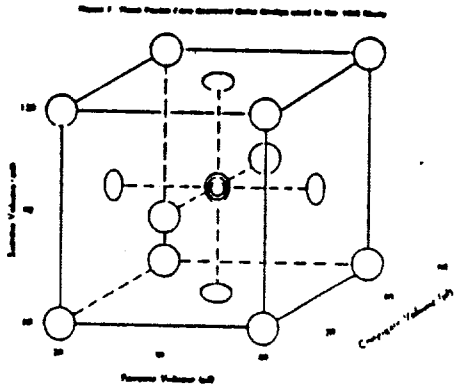


Multidimensional point plots are most common graphical DOX tool.

Examples here show factorial point plots for one-, two-, three-, and four-factor designs.

Source: Andrews (1964).

10



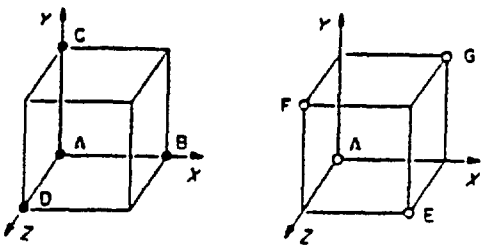
Another application, some minor variations in presentation form. Dashed lines help locate face-centered points.

Source: Myers (1985).

VIEWGRAPH

TEXT

11



Youden: No scientist, when presented with these designs graphically, would prefer the one-at-a-time version on the left.

Aesthetic property apparent here: span the design space. Will return to this again later.

Source: Youden (1972).

12

TABLE 4.

Program for three variables, two with three classes, one with two classes.

Variables		
x	z	X
y	y	Y
z		Z

Six sets					
1	2	3	4	5	6
x	x	x	x	X	X
y	y	y	Y	y	Y
z	Z	Z	z	z	Z

	z	z	X
y	1 2	Z −2	
y	Z 1		1 −1
Y		1 −1	Z 1

Above coefficients are weighing factors to estimate $z - z'$

Youden's approach to representing an incomplete design, circa 1962:

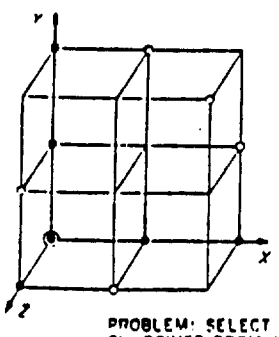
A TABLE

Source: Youden (1962).

VIEWGRAPH

TEXT

13



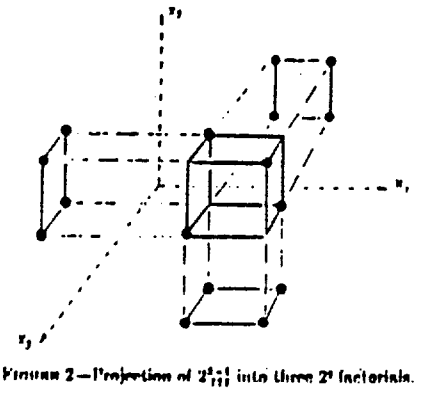
Youden's choice for representing the same design, circa 1972:

A PLOT

Plot gives visual hints to confounding pattern that can be used not just to display designs, but to create them as well.

Source: Youden (1972).

14



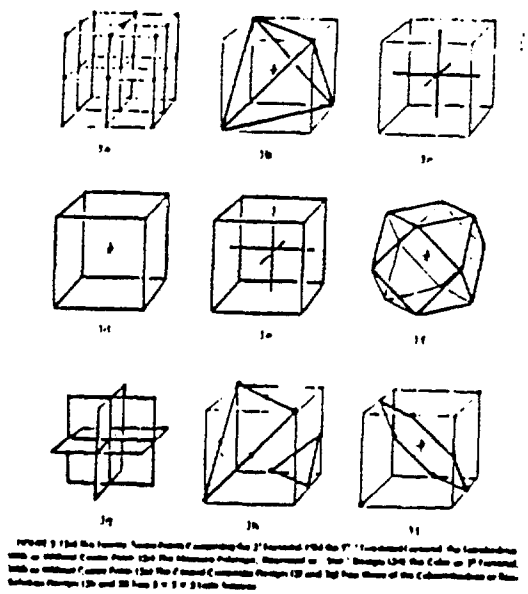
Box and Hunter used graphical models of designs, and studied their projections to find ones with "balance".

Source: Box and Hunter (1961).

VIEWGRAPH

TEXT

15

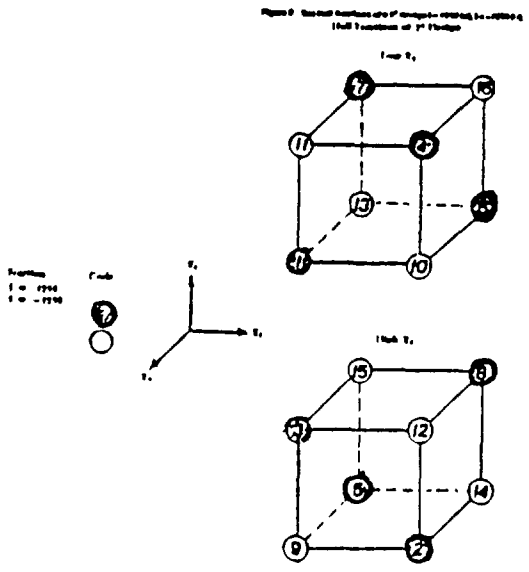


Hunter used multidimensional point plots (with appropriate reference lines) to illustrate many common designs.

Here, graphical representation is for analytical use -- the designs had already been created.

Source: Hunter (1985).

16



Plots used effectively to illustrate fractional design for an industrial application. This and two following viewgraphs show fractional factorial plots from recently published applications.

Shading here used to identify each of the two half-fractions.

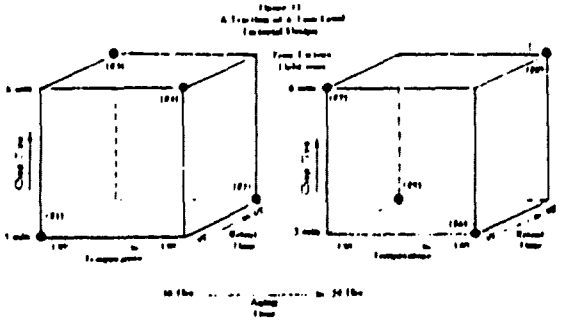
Run order shown inside bubbles.

Source: Snee (1985a).

VIEWGRAPH

TEXT

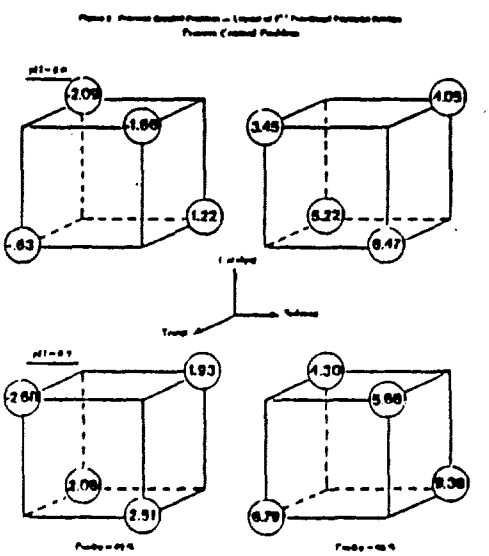
17



Here the actual factor levels are used to label a 2^{4-1} design.

Source: Andrews (1964).

18



Developing and understanding a graphical representation for the design can later be augmented to display the results of the experiment.

Bubbles of this 2^{5-1} design show outcomes of experiments.

Extension: use a symbol that conveys both location AND spread at each design point when design includes replication (or is an inner-outer design a-la Taguchi).

Source: Snee (1985a).

So far, shown designs displayed graphically to reveal properties. That is, plots used DESCRIPTIVELY. How to use graphical methods to GENERATE DESIGNS for particular applications? READ ON ---------->

(most graphical references use plots for analysis or presentation, not for design generation)

VIEWGRAPH

TEXT

19

"...It is proved (Appendix 1) that if a polynomial of *any* degree d_1 is fitted by the method of least squares over *any* region of interest R in the k variables, when the true function is a polynomial of any degree $d_2 > d_1$, then the bias averaged over R is minimized for all values of the coefficients of the neglected terms, by making the moments of order d_1+d_2 and less of the design points equal to the corresponding moments of a uniform distribution over R."

- G.E.P. Box and N.R. Draper

HOW TO GENERATE DESIGNS GRAPHICALLY:

PRINCIPLE #1
<----------

(i.e. spread points out uniformly over space)

Source: Box and Draper (1959).

20

".. convenient to regard designs as built up from a number of component sets of points, each set having its points equidistant from the origin ..."

"... form the vertices of a regular polygon, polyhedron, or polytope..."

- Box and Hunter (1957)

HOW TO GENERATE DESIGNS GRAPHICALLY

PRINCIPLE #2
<----------

(if whole design too complex, use divide-and-conquer strategy to design smaller components to be combined -- see viewgraphs 30 and 31)

Source: Box and Hunter (1957).

VIEWGRAPH

TEXT

21

"Choose new points to MAXIMIZE the minimum distance from all existing design points.."

-Kennard and Stone (1969)

HOW TO GENERATE DESIGNS GRAPHICALLY

PRINCIPLE #3
<----------

(this consideration arises from "optimal" design considerations -- min variance for first order model terms)

Source: Kennard and Stone (1969).

22

SOME USEFUL CONCEPTS
for generating
GOOD DESIGNS
from
MULTIDIMENSIONAL POINT PLOTS

1 COVER THE DESIGN SPACE UNIFORMLY

2 DECOMPOSE COMPLICATED DESIGNS INTO GRAPHICAL SUBCOMPONENTS

3 SPAN THE WHOLE DESIGN SPACE:
MAKE ADDED DESIGN POINTS FAR FROM EXISTING POINTS TO MINIMIZE VARIANCE FOR FIRST ORDER EFFECTS

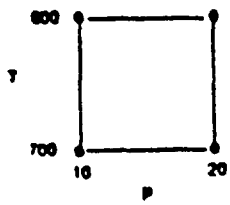
4 CHECK PROJECTIONS TO PLANES AND LINES

Last point, used extensively by Box and Hunter, was mentioned earlier.

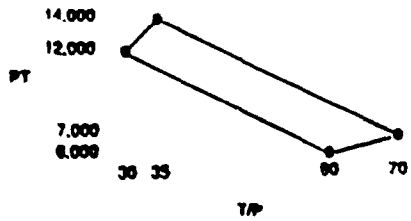
VIEWGRAPH

TEXT

23



$y = a_0 + a_1(T) + a_2(P) + \varepsilon \Rightarrow$ BALANCED



$y = a_0 + a_1(PT) + a_2(T/P) + \varepsilon \Rightarrow$ UNBALANCED

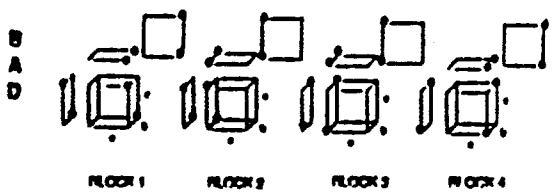
Box and Draper findings above give some model independence.

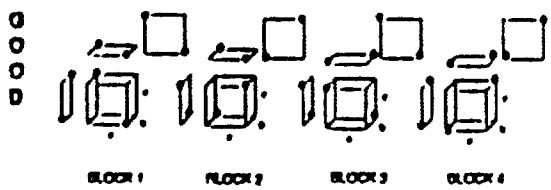
BUT using graphical methods to generate designs does not free us from the fact:

DESIGN GOODNESS DEPENDS ON THE TRUE FORM OF THE MODEL BEING INVESTIGATED.

Source: Satterthwaite (1959).

24

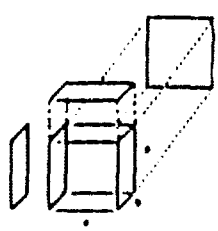




BLOCKED DESIGNS FROM

BOX, HUNTER & HUNTER,

(pp339-341)



To illustrate multidim. point plots for design, first show a 3-factor experiment to be run in 4 blocks of 2.

Decomposition, projection, and spanning (points 2, 3, &4) used to generate good design here. (Decomposition is of cube points into 4 sets of antipodal pairs).

Block effects confounded with main effects in bad design seen from top and rear projections.

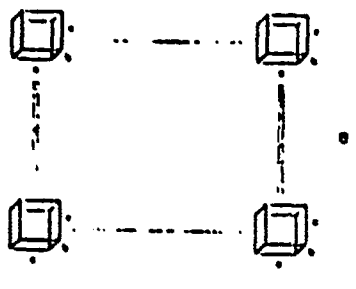
The relative merits of these two designs much easier to see here than in their original (non-graphical) description.

Source: Box, Hunter, and Hunter (1978).

VIEWGRAPH

TEXT

25

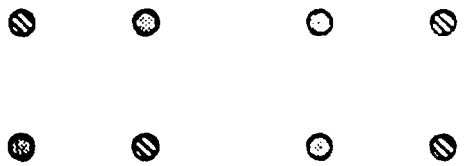


A 2^5 Factorial Design

(w/ 4-block layout: factors to factors, blocks to blocks)

This figure shows a multidim. point plot for a 2^5 factorial design. The blocks for the design above (blocked designs from BH&H) set in a row rather than a square because additional structure here not present above.

Example: block effect (1+4)-(2+3) would appear as an "interaction" pattern here, while an equivalent pattern, (1+2)-(3+4) would have a main effect pattern.

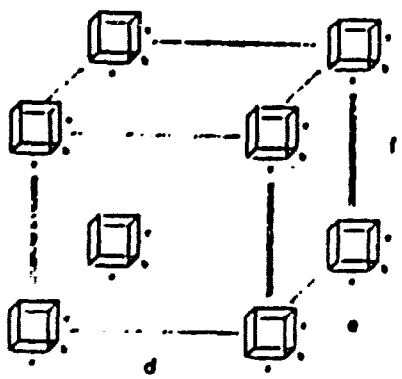


INTERACTION PATTERN

MAIN EFFECT PATTERN

These plots and those to follow are easy to generate and manipulate using a Macintosh (MacDraw ©). Pro- jections are NOT automatic, though.

26

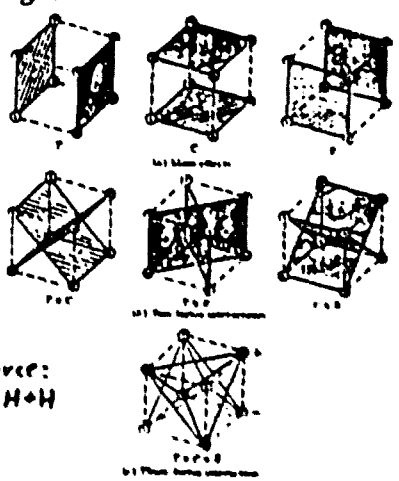


A 2^6 Factorial Design

Figure at left can be used as a template for designs with 7 or more factors.

Fill in subset of dots at small cube vertices to generate an incomplete or fractional 2^6 design.

Use two dot symbols, e.g. ● ■ for a 2^7 design.

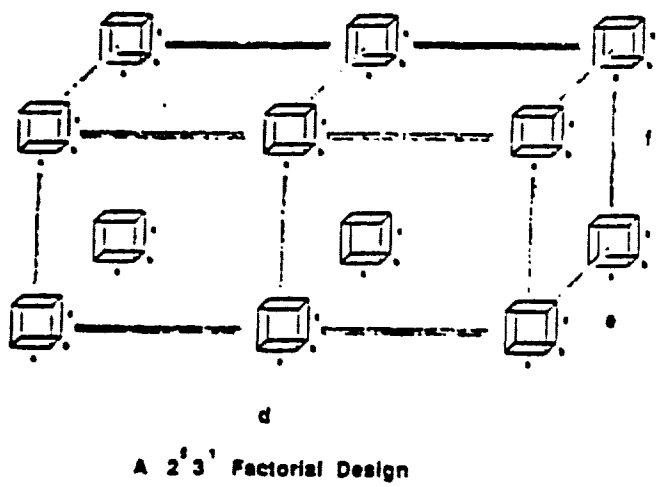


Source: B,H+H

VIEWGRAPH

TEXT

27



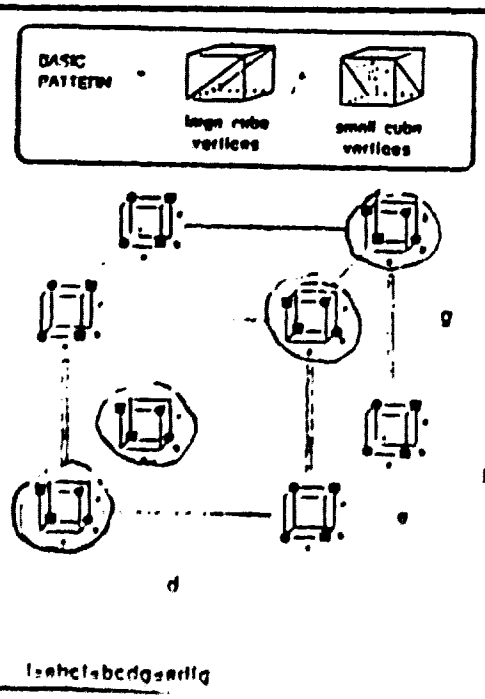
A $2^2 3^1$ Factorial Design

Can do multidimensional point plots for $2^m 3^n$ designs, too.

Compare with Youden plot earlier, and viewgraphs 31-34.

2-way interaction patterns

28



Second example from literature is a 2^{7-2} fractional factorial. Next three viewgraphs illustrate the three designs presented in the reference.

Value of "minimum aberration" designs is consonant with graphical design principles.

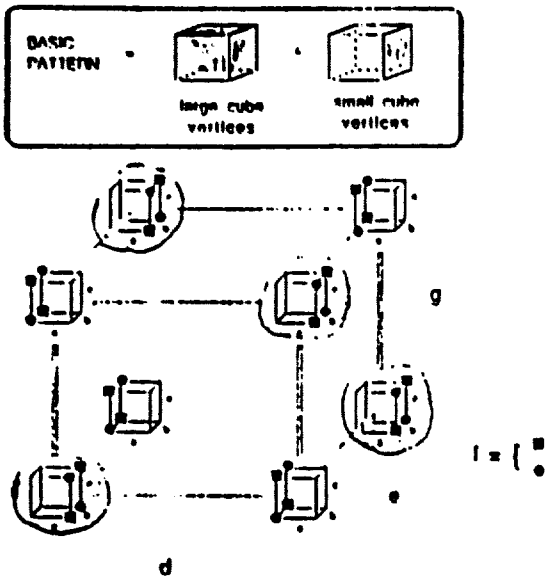
For 2^k designs, use decomposition and idea that best fractions span the space: best point allocation, therefore, is based on three way interaction pattern.

Reference: Fries and Hunter (1980).

VIEWGRAPH

TEXT

29



Identical small-cube forms denoted by circles.

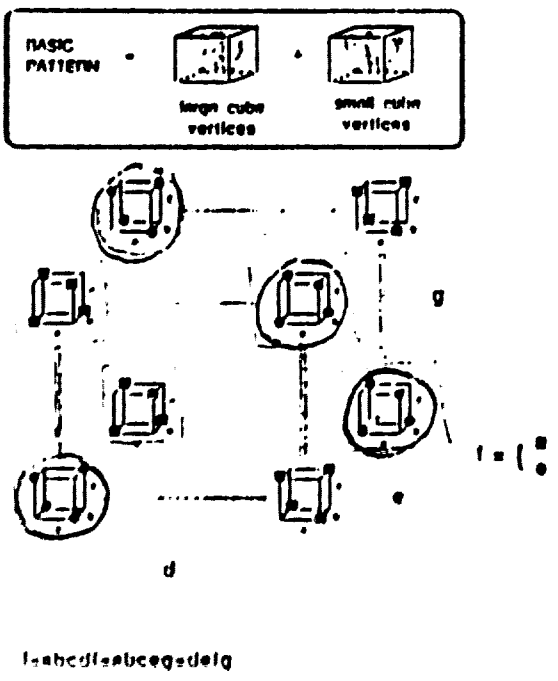
Good large-cube pattern.

Poor small-cube pattern - can be fixed.

All projections can be visualized without much trouble.

Reference: Fries and Hunter (1980).

30



THE MINIMUM ABERRATION DESIGN

Pattern here is good; still some flaws - the choice of the particular 2^{7-4} small-cube pattern has a 2-way pattern on the large cube, and two way pattern separates levels of f based on levels of d (I=defg).

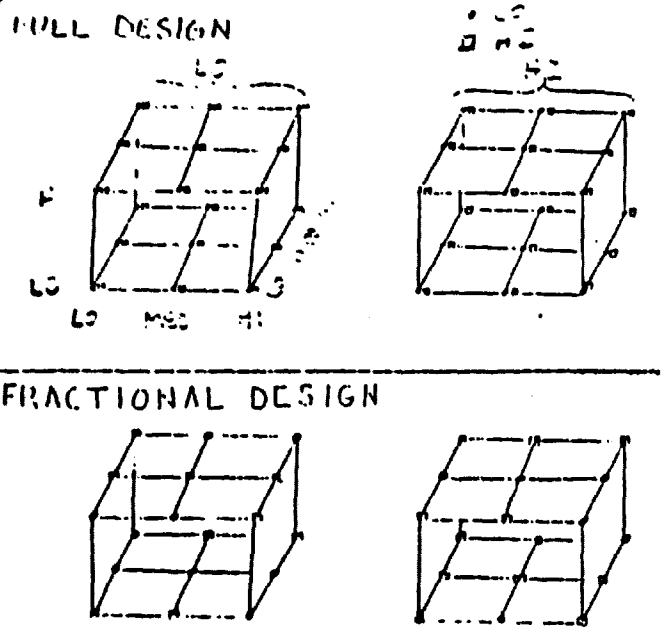
At this point, can only push confounding around; not enough design points to fix.

Reference: Fries and Hunter (1980).

VIEWGRAPH

TEXT

31



Minimum aberration and incomplete block examples were from academic literature.

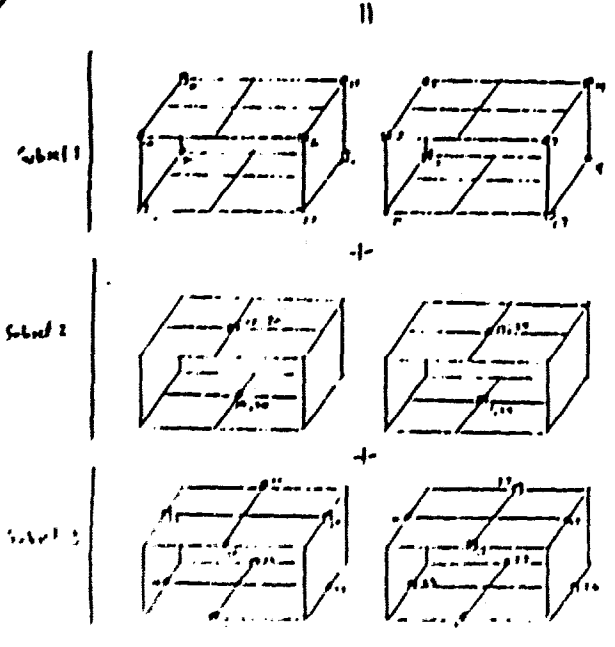
This example from RCA, industrial research problem. Design was generated graphically, as shown here, for an experiment in 1982.

Full factorial was a $2^3 \quad 3^2$.

Designed a 1/2 fraction.

Source: Barton (1982).

32



The 1/2 fraction was composed of three pieces, following DESIGN PRINCIPLE #2. Easy to see (and to design) this way.

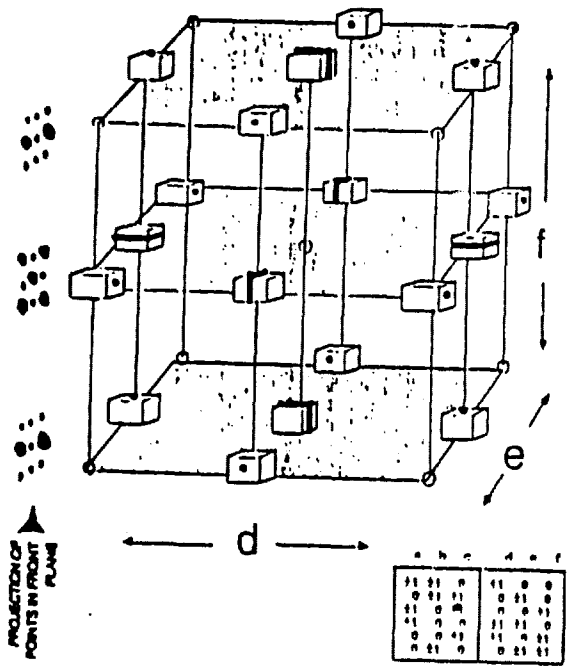
Note: numbers represent run order, which was modified in final design.

Source: Barton (1982).

VIEWGRAPH

TEXT

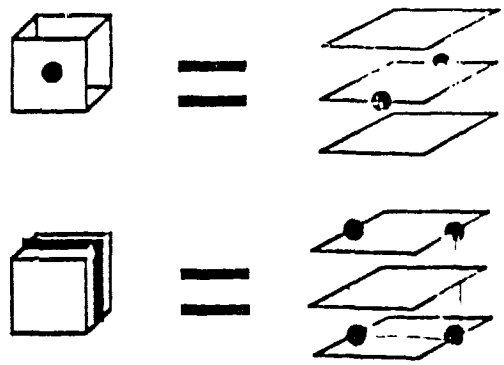
33



Box - Behnken 3^4 Fractional Design

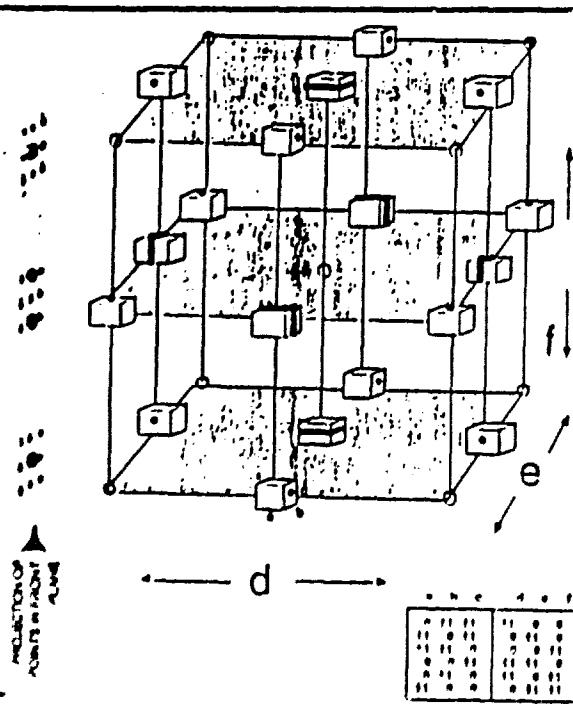
BOX-BEHNKEN DESIGN

Illustrates use of icons for complicated multi-dimensional point plots:



Reference: Box and Behnken (1960).

34



Bad - Barton 3^4 Fractional Design

Easy to generate alternative fractions using the icons; Bad Barton at left.

Some properties of both designs immediately obvious:

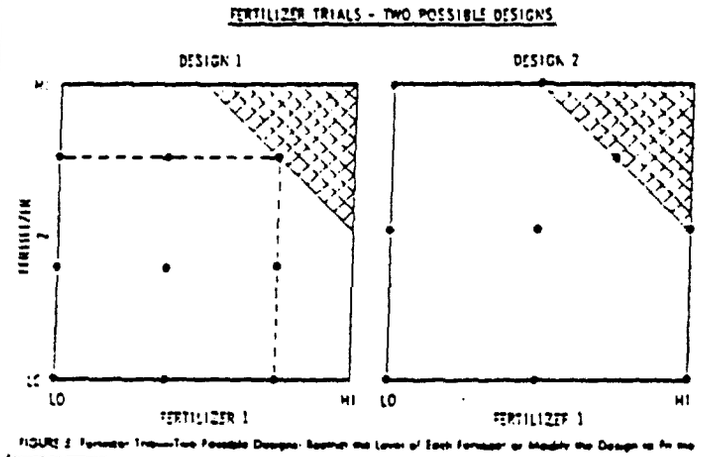
no center
no extreme vertices (violates #3)

Other properties (like why Bad-Barton is bad) not obvious without projections.

VIEWGRAPH

TEXT

35

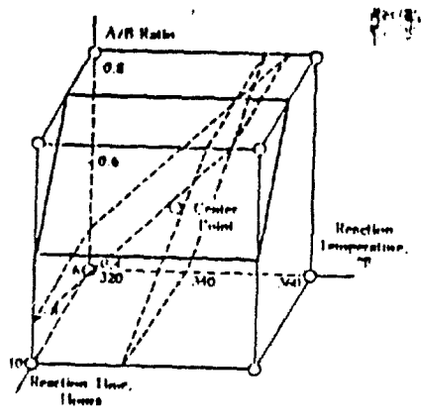


Multidimensional point plots for factorial designs allow intuitive modifications to incorporate constraints on the design space.

Snee (1981) gives rules used by CONSIM to place mixture design points on boundaries caused by constraints. First example in this presentation of "mental graphics".

Source: Snee (1985b).

36



For many practical problems, constraints are few enough to allow visualization-

-and better control of the design.

Source: Kinzer (1985).

VIEWGRAPH

TEXT

37

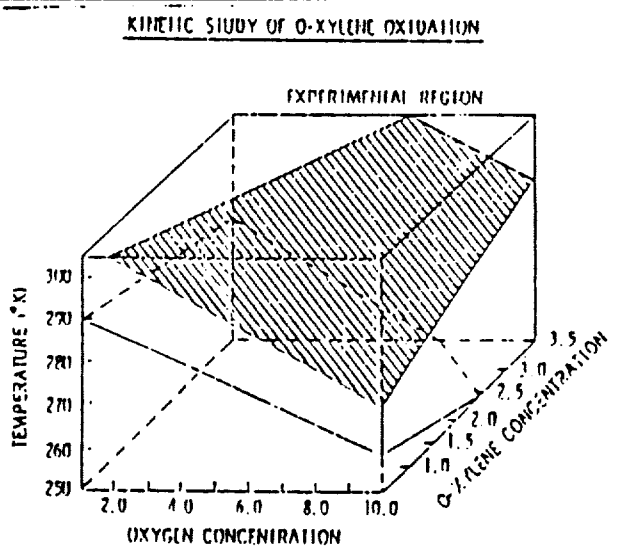


FIGURE 6. Ortho-Xylene Oxidation Kinetic Study—Experimental Region is Defined by Two Nonparallel Planes in Three Dimensions (Jousola, Bacon, and Downie (1972)).

Another example showing constraints limiting the experimental region.

Note: complex constraints may suggest a transformation to the model factors.

Source: Snee (1985b).

38

FACTORIAL MULTIDIMENSIONAL PLOTS

-

SUMMARY

Multidim. point plots are useful concepts even when they can't actually be drawn. Fry [] uses "mental graphics" to construct fractional 2 3 designs from hypersphere designs composed of multiple sets of 2 designs.

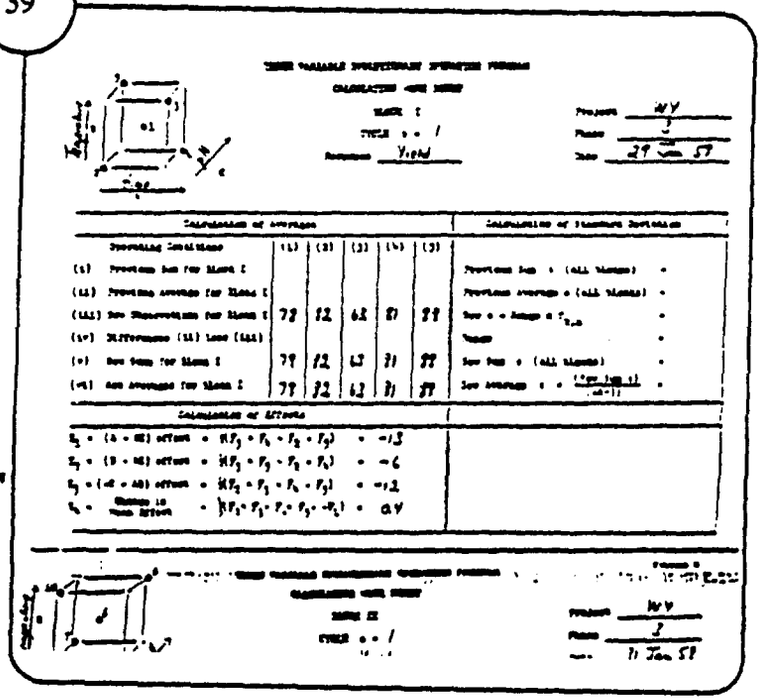
Why factorial (hypercube)?
answer: limits # of factor levels, easier to do math, plot results, and view design in 2-D, 3-D, etc.

NEXT SECTION REVIEWS SPECIAL METHODS FOR RESP. SURFACE / EVOP

VIEWGRAPH

TEXT

39

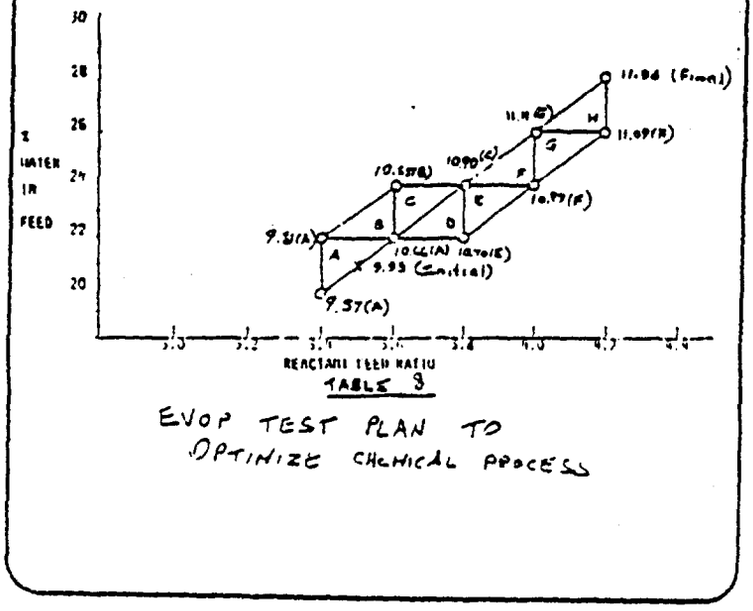


In an early EVOP worksheet, multidim. point plots for design were part of the data collection worksheet.

Graphical design provides layout to run the experiment from.

Source: Box and Hunter (1959).

40



A simplex plan that is updated as runs are completed can be used to choose the next run point.

This is graphical sequential design.

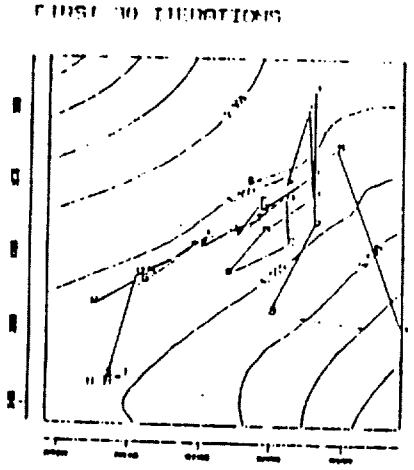
Easier if superimpose contours of model fitting a recent subset of observations; see next viewgraph.

Source: Hahn,Bemesderfer,and Olsson (1986)

VIEWGRAPH

TEXT

41



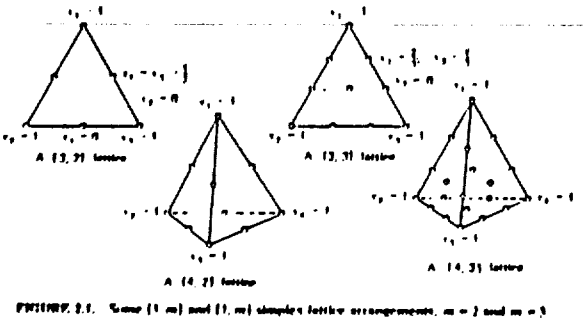
Optimization trajectory superimposed on the function value surface

Here the model is not a polynomial in the usual (Taylor approximation) sense, but Hardy's [] interpolation function.

Source: Barton (1985).

Reference: Hardy (1971).

42



Above trajectory was for a Nelder-Mead simplex sequential optimization strategy.

Here are simplices of a different sort for DOX: simplices arising from mixture experiments.

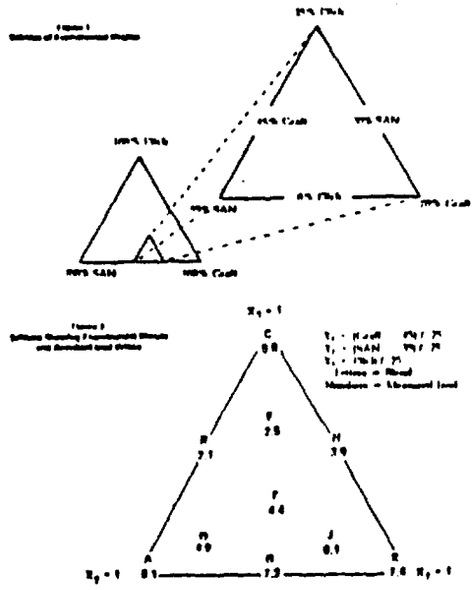
The next few viewgraphs review graphical representations that have been used to create and analyze mixture designs.

Source: Cornell (1981).

VIEWGRAPH

TEXT

43



As for factorial designs, point plots can be used to identify subregions for study.

In addition to the usual mixture constraint, most real mixture problems have additional requirements that limit the design space.

Source: Koons and Wilt (1985).

44

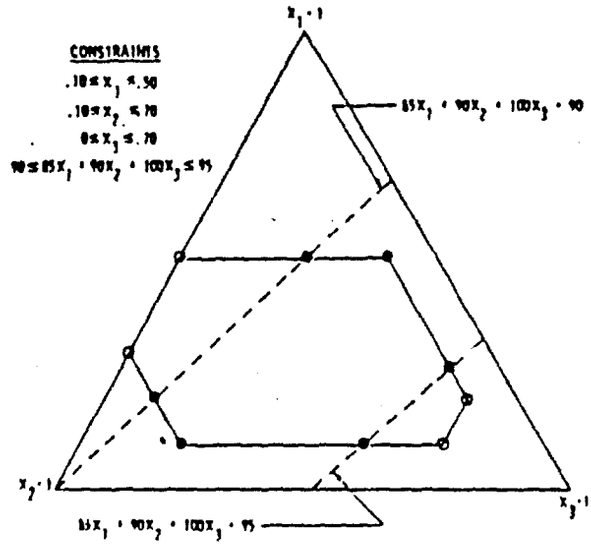


Figure 4. *Three-Component Mixture System With Single Component and Multiple Component Constraints*

More complicated constraints yield irregularly shaped regions.

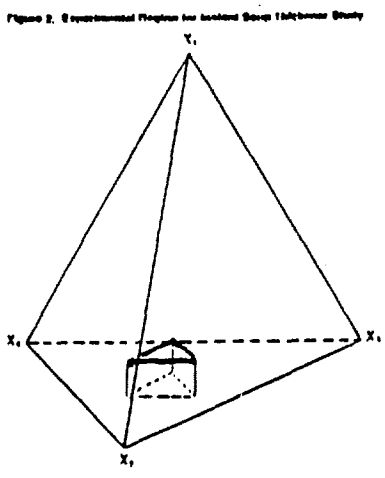
Snee's XVERT program depends on the geometric concepts of edges, vertices, and face centroids to select "good" design points. Again, this is "mental graphics", since a graphical image is used, but it is not actually drawn.

Source: Snee (1981).

VIEWGRAPH

TEXT

45



Four factor mixture experiments and constrained subsets can be drawn effectively, and have been used in industry.

Source: Hare (1985).

46

NOMOGRAMS

and other

GRAPH PAPER GRAPHS

This ends material on multidimensional point plots for DOX.

Nomograms and graph paper graphs are practical tools for DOX, but they are not in the spirit of earlier material. Only a brief sample here to illustrate the kind of advantages they offer.

VIEWGRAPH

TEXT

47

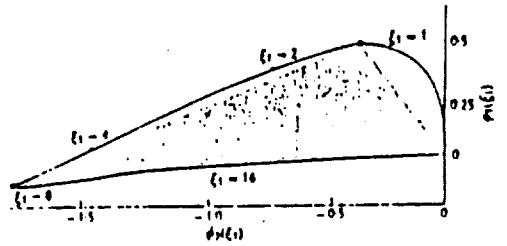


Fig. 2. The design locus for the function $f(\xi_1; 0.7, 0.2)$ with the optimal solution indicated by heavy dots.

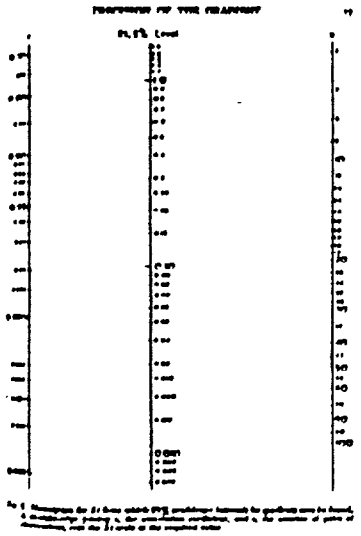
A graphical aid for D-optimal design

(Box & Lucas, 1959)

Graphical technique here is one step removed from design. It represents a mathematical function of the design structure.

Source: Box and Lucas (1959).

48



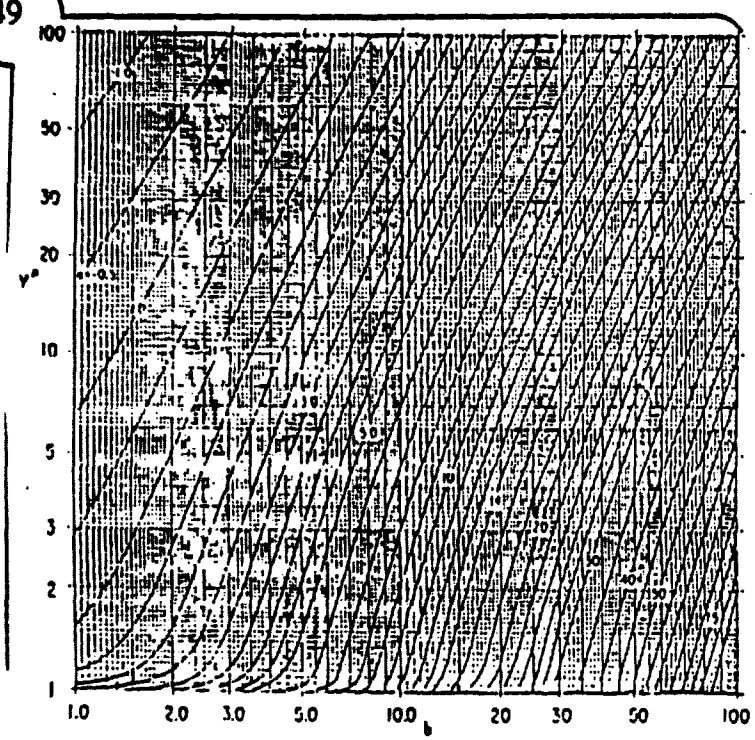
This nomogram allows experimenter to choose sample size required for desired accuracy of the slope coefficient.

Source: Beech

VIEWGRAPH

TEXT

49

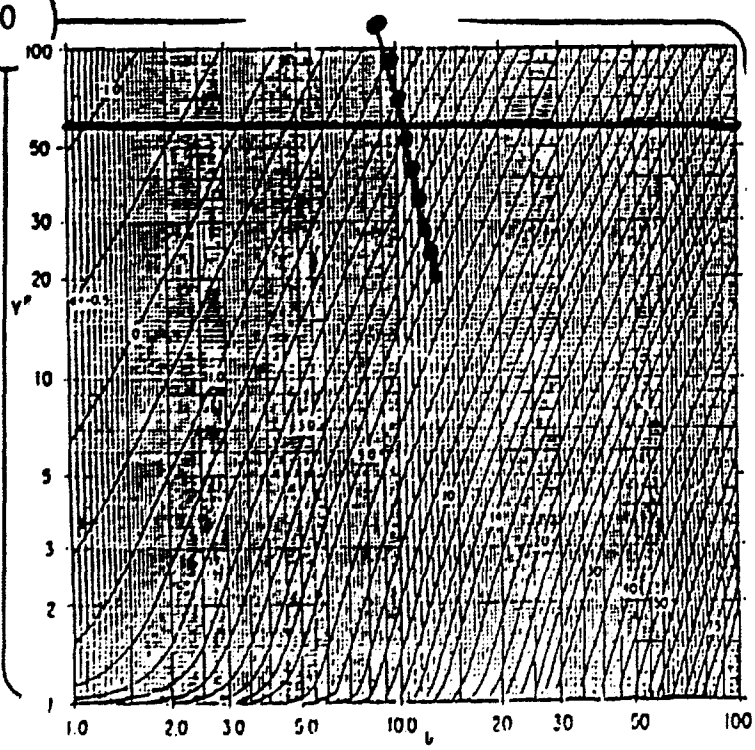


Like a nomogram, this graph is used to show the variance of maximum likelihood estimates as a function of design parameters.

The model here is Arrhenius; design parameters are test temperatures and test time. Censored observations are expected.

Source: Nelson and Kielpinski (1975).

50



Because design properties are displayed graphically, it is possible to optimize other design properties (i.e. other than variance of estimates) by making graphical additions!

Example: minimize the maximum test temperature without exceeding a variance limit.

Source: Barton and Nelson (1987).

VIEWGRAPH

TEXT

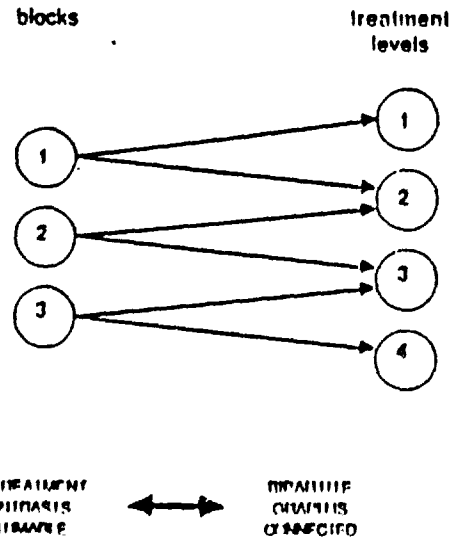
51

NETWORK
DESIGN
REPRESENTATIONS

Like the nomograms and graph-paper graphs, the network design representations to follow are one level removed from the design.

Because of this, expect that they will be less useful for design synthesis.

52



These plots, due to Butz, relate connectivity to estimable contrasts.

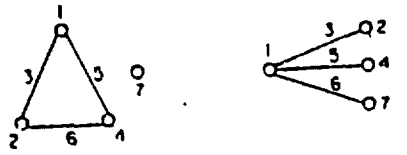
For small examples, these plots can be used to set up and evaluate designs for ANOVA models.

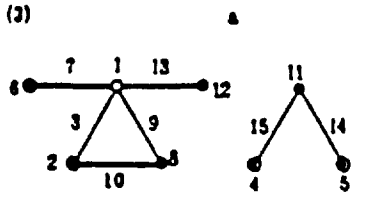
Source: Butz (1982).

VIEWGRAPH

TEXT

53



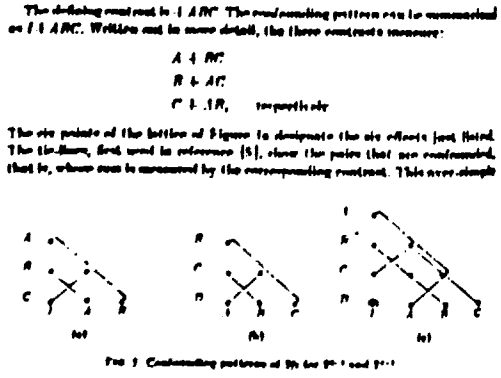


Taguchi uses "linear graphs" to expose confounding patterns in fractional designs. They appear useful for choosing a defining relation that yields a desired confounding pattern.

Method of construction: unknown

Source: Taguchi (1980).

54



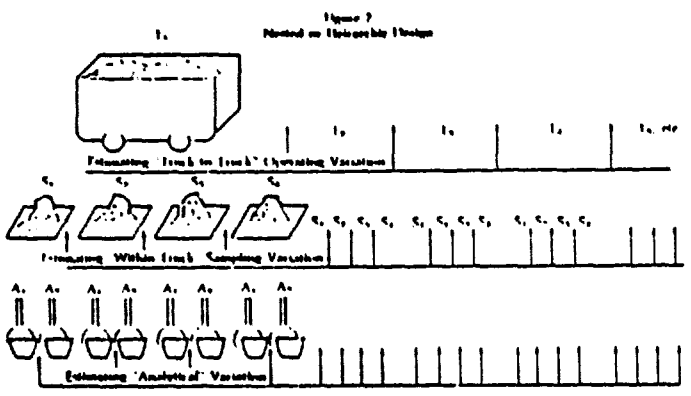
Cuthbert Daniel's method for displaying confounding patterns is more difficult to see (for me). Used to analyze rather than generate.

Source: Daniel (1962).

VIEWGRAPH

TEXT

55

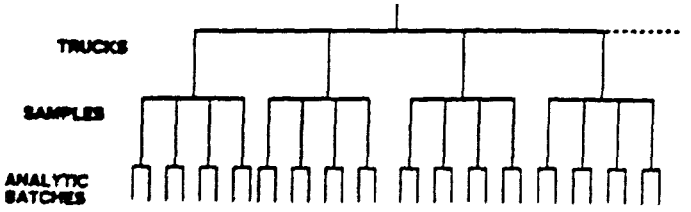


Graphical representations of hierarchy help to develop nested designs for mixed and random effects models.

Andrews was particularly graphic.

Source: Andrews (1964).

56



SCHEMATIC for ANDREWS' HIERARCHIC DESIGN

A simpler, perhaps less informative representation of the same design. This form has been used by several authors.

See also: Leone,Nelson,and Johnson (1968)

VIEWGRAPH

TEXT

57

What graphical concepts do these tools exploit?

1 Design Balance/Symmetry
2 Design Projections
3 "Face" Incidence of Design Points
4 Network properties: connectedness, etc.
5 Analog Computations

58

What are the strengths and limitations of graphical methods?

+ Flexible
  - make tradeoffs visually
  - incorporate constraints graphically
+ Robust
+ Uses powerful computer - human eye
+ Graphical DOX methods easy to use & remember

- Non-quantitative
- Dimensional limitations

VIEWGRAPH

TEXT

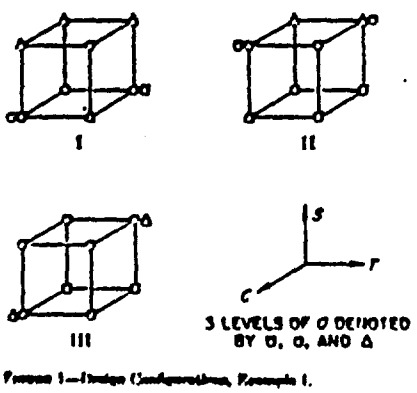
59

What role can computers play in graphical DOX?

1. Make descriptive tools into prescriptive ones
   - rapid plotting of alternative designs
   - exhaustive plots of alternative designs for scanning
2. Interactive graphics
   - real time design manipulation
   - computed design properties updated and displayed
3. Rule-based systems to manipulate geometric or network objects

Of course, computers play other roles in DOX, e.g. DETMAX. Here we mean getting computers to help with the plotting, projections, views, etc.

60



Even for DETMAX applications, graphical methods resorted to for understanding and evaluation.

Source: Mitchell (1974).

VIEWGRAPH

TEXT

61(end)

Summary – the place of graphical methods in DOX

1 Graphical: ***Investigative, creative***

2 Mathematical, Computer-Aided: ***confirmatory***

Examples of graphics shown here aren't meant to be prescriptive; graphical DOX as a distinct entity is too new.

This selection represents useful methods to trigger your own imagination.

Try to find useful ways to handle designs with many factors.

USE YOUR RIGHT BRAIN
(and may the force be with you!)

Reference: Box (1984).

NEXT:

GRAPHICAL ANALYSIS

Andrews, H.P. (1964), "The Role of Statistics in Setting Food Specifications," *Proceedings of the Sixteenth Research Conference of the Research Council of the American Meat Institute*, 43-56. Reprinted in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control, 1985.

Barton, R.R. (1982). RCA Internal Correspondence to I. Chen.

Barton, R.R. (1985), RCA Internal Report on CONTURIT.

Barton, R.R., and Nelson, W.B. (1987), "Optimum Accelerated Lifetest Plans which Minimize the Maximum Test Stress," (in preparation).

Beech, D.G. (1961), "Some Notes on the Precision of the Gradient of an Estimated Straight Line," *Applied Statistics, 10, 14-31.*

Box, G.E.P. (1954), "The Exploration and Exploitation of Response Surfaces: Some General Coslderations and Examples," *Biometrics*, 10, 16-60.

Box, G.E.P. (1984), "The Importance of Practice in the Development of Statistics," *Technometrics*, 26, 1-8.

Box, G.E.P., and Behnken, D.W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics*, 2, 455-475.

Box, G.E.P., and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-653.

Box, G.E.P., and Hunter, J.S. (1957), "Multifactor Experimental Designs for Exploring Response Surfaces," *Annals of Mathematical Statistics*, 28,195-241.

Box, G.E.P., and Hunter, J.S. (1959), "Condensed Calculations for Evolutionary Operation Programs," *Technometrics*, 1, 77-95.

Box, G.E.P., and Hunter, J.S. (1961), "The 2^{k-p} Fractional Factorial Designs Part I," *Technometrics*, 3, 311-351.

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978), *Statistics for Experimenters*, New York: Wiley.

Box, G.E.P., and Lucas, H.L. (1959), "Design of Experiments in Nonlinear Situations," *Biometrika*, 46, 77-90.

Box, G.E.P., and Wilson, K.B. (1951), "On the Experimental Attainment of Optimal Conditions," *Journal of the Royal Statistical Society series B*, 13, 1-38.

Butz, L. (1982), "Connectivity in Multi-Factor Designs: A Combinatorial Approach," *Research and Education in Mathematics 3*, Berlin: Helderman Verlag.

Cornell, J.A. (1981), *Experiments with Mixtures*, New York: Wiley.

Daniel, C. (1962), "Sequences of Fractional Replicates in the 2^{p-q} Series," *Journal of the American Statistical Association*, 58, 403-429.

DeBaun, R.M. (1959), "Response Surface Designs for Three Factors at Three Levels," *Technometrics*, 1, 1-8.

Feder, P.I. (1974), "Graphical Techniques in Statistical Data Analysis," *Technometrics*, 16, 287-319.

Fries, A. and Hunter, W.G. (1980), "Minimum Aberration 2^{k-p} Designs," *Technometrics*, 22, 601-608.

Fry, K.E. (1961), "Finding New Fractions of Factorial Experimental Designs," *Technometrics*, 3, 359-370.

Hahn, G.J. (1984), "Experimental Design in the Complex World," *Technometrics*, 26, 19-31.

Hahn, G.J., Bemesderfer, J. and Olsson, D.M. (1986), "Explaining Experimental Design Fundamentals to Engineers: A Modern Approach," unpublished.

Hardy, R.L. (1971), "Multiquadric Equations of Topography and Other Irregular Surfaces," *Journal of Geophysical Research*, 76, 1905-1915.

Hare, L.B. (1985), "Graphical Display of the Results of Mixture Experiments," in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Hunter, J.Stuart (1985), "Statistical Design Applied to Product Design," *Journal of Quality Technology*, 17, 210-221.

Ishikawa, K. (1982), *Guide to Quality Control*, Tokyo: Asian Productivity Organization, (second edition).

John, J.A. and Mitchell, T.J. (1977), "Optimal Incomplete Block Designs," *Journal of the Royal Statistical Society, Series B*, 39, 39-43.

Kennard, R.W., and Stone, L.A. (1969), "Computer Aided Design of Experiments," *Technometrics*, 11, 137-148.

Kinzer, G.R. (1985), "Application of Two-Cubed Factorial Designs to Process Studies," In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Koons, G.F., and Wilt, M.H. (1985), "Design and Analysis of an ABS Pipe Compound Experiment," In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Leone, F.C., Nelson, L.S., and Johnson, N.L. (1968), "Sampling Distributions of Variance Components: Empirical Studies of Unbalanced Nested Designs," *Technometrics*, 10, 719-726.

Mitchell, T.J. (1974), "An Algorithm for the Construction of D-optimal Experimental Designs," *Technometrics*, 16, 203-210.

Myers, G.C. Jr. (1985), "Use of Response Surface Methodology in Clinical Chemistry," In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Nelson, W.B., and Kielpinski, T.J. (1975), "Optimum Accelerated Lifetests for the Normal and Lognormal Life Distributions," *IEEE Transactions on Reliability*, 24, 310-320.

Raktoe, B.L., Heydayat, A., and Federer, W.T. (1981), *Factorial Designs*, New York: Wiley.

Satterthwaite, F.E. (1959), "Random Balance Experimentation," *Technometrics*, 1, 111-137.

Snee, R.D. (1981), "Developing Blending Models for Gasoline and Other Mixtures," *Technometrics*, 23, 119-130.

Snee, R.D. (1985a), "Experimenting with a large Number of Variables," In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R.D. Snee, L.B. Hare, and J.R. Trout, Milwaukee: American Society for Quality Control.

Snee, R.D. (1985b), "Computer-Aided Design of Experiments-- Some Practical Experiences," *Journal of Quality Technology*, 17, 222-236.

Taguchi, G., and Wu, Y. (1980), *Introduction to Off-Line Quality Control*, The Central Japanese Quality Control Association (available from American Supplier Institute, 32100 Detroit Industrial Expressway, Romulus, MI 48174.

Youden, W.J. (1962), "Systematic Errors in Physical Constants," *Technometrics*, 4, 111-123.

Youden, W.J. (1972), "Enduring Values," *Technometrics*, 14, 1-10.

DESIGN OF EXPERIMENTS CONFERENCE
16-20 OCTOBER 1989
ATTENDEE LIST

2 OCT 89

NAME	SPOUSE	TUTOR/ CONF.	CONF.	BANQ.	RESV HOTEL
DR. EUGENE F. DUTOIT		X	X	X	X
PROF. ROBERT BECHHOFER			X	X	X
PROF. SHELEMYALN ZACKS			X	X	X
DR. MALCOLM S. TAYLOR		X	X	X	X
LARRY W. BRANTLEY		X	X	*	X
PROF. D.L. TAYLOR			X	*	X
SHIRISH CHINCHALKAR			X	*	X
DR. NOZER D. SINGPURWALLA			X	*	X
J. ROBERT BURGE		X	X	X	X
ROBERT BURTON		X	X	X	X
LTC LEE S. DEWALD SR.	X		X	X	NPGS
DR. DAVID F. CRUESS	*	X	X	X	X
CARL B. BATES			X	*	X
MARK A. YOUNGREN			X	*	X
RICHARD M. BRUGGER			X	*	X
GARY AASHEIM			X	*	X
DONALD L. INGLEHART			X	X	X
W. J. CONOVER	*		X	X	X
DR. RON HOCKING		X	X	*	X
DR. GERALD R. ANDERSEN	X	X	X	X	X
DR. FRANCIS G. DRESSEL	X	X	X	X	X
ANN E.M. BRODEEN				X	X
THOMAS MATHEW			X	X	DAYS INN
BIMAL K. SINHA			X	X	DAYS INN
DAVID W. SCOTT			X	X	X
DR. GARY G. KOCH			X	*	X
RUSSELL R. BARTON			X	X	X
DR. IZHAK RUBIN			X	X	X
WILLIAM BAKER			X	X	X
DR. BARRY BOOT		X	X	X	X
LTC RICKEY A. KOLB	X	*	X	X	X
DR. EMANUEL PARZEN			X	X	X
LARRY G. HAYNES		X	X	*	X
CARL T. RUSSELL		X	X	*	X
RICHARD A. KASS		X	X	X	X
WILLIAM AGEE	**		X	X	X
DEAN TAYLOR			X	X	X
DR. ANTHONY TRIPPE		X	X	X	X
DR. MELVIN BROWN		X	X	*	X
PROF. JAYARAM SETHURAMAN	X		X	X	X
DR. FLOYD JOHN		X	X	X	X
CARL PFEIFFER	**		X	*	X
ED GUSSIE			X	*	X
CHARLES CORREIA		X	X	*	X
DELORIS TESTERMAN			X	*	X
ERNEST SCHERB			X	*	X
DR. DOUGLAS TANG		X	X	X	X
DR. JOCK O. GRYNOVICKI			X	X	X
DONALD E. McCOY	**	X	X	X	NPGS
DR. JIM MAAR		X	X	X	X
DR. BERNARD HARRIS	*	*	X	X	X
MIRCEA GRIGORIU		X	X	*	X

* DID NOT SPECIFY ON TEC RESPONSE FORM
** TWO INDIVIDUALS IN PARTY-DID NOT IDENTIFY

+This list of attendees was made out before the conference started. Drs. R, Bechhofer, B.Harris, D.L. Inglehart, and N.D. Singpurwalla, and maybe others were unable to attend because of the earthquake.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp. Date Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARO REPORT 90-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Army Research Office 6b. OFFICE SYMBOL (If applicable) SLCRO-ARO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 12211 Research Triangle Park, NC 27709	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AMSC on behalf of ASO (ADA) 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Proceedings of the Thirty-Fifth Conference on the Design of Experiments in Army Research, Development and Testing

12. PERSONAL AUTHOR(S)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM Jan 89 TO Feb 90	1990 August	327

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This is a technical report of the Thirty-Fifth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at this meeting. These articles treat various Army statistical and design problems.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Francis G. Dressel	22b TELEPHONE (Include Area Code) 919-549-0641, ext. 324 22c OFFICE SYMBOL SLCRO-MA

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE