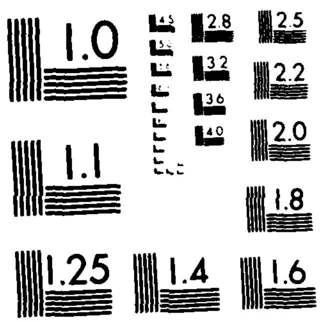
AD-A136 567 EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES 1/1
IF THEY ARE ACTUA..(U) OHIO STATE UNIV RESEARCH
FOUNDATION COLUMBUS M L MOESCHBERGER ET AL. 26 OCT 83
UNCLASSIFIED AFOSR-TR-83-1278 AFOSR-82-0307 F/G 12/1 NL

END
DATE
FILMED
2 84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AFOSR-TR- 83-1278

(4)

RF Project 763265/714837
Annual Report

EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES, IF THEY ARE ACTUALLY DEPENDENT, IN A SERIES SYSTEM

Melvin L. Moeschberger
Department of Preventive Medicine

and

John P. Klein
Department of Statistics

A136567

For the Period
September 1, 1982 - September 30, 1983

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, D.C. 20332

Grant No. AFOSR 82-0307

October 26, 1983

OSU
The Ohio State University
Research Foundation
1314 Kinnear Road
Columbus, Ohio 43212

84 01 04 098

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 83-1278

6a. NAME OF PERFORMING ORGANIZATION
Ohio State University

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Air Force Office of Scientific Research

6c. ADDRESS *(City, State and ZIP Code)*
Department of Preventive Medicine
and Department of Statistics
Columbus OH 43212

7b. ADDRESS *(City, State and ZIP Code)*
Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL *(If applicable)*
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-82-0307

8c. ADDRESS *(City, State and ZIP Code)*
Bolling AFB DC 20332

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE *(Include Security Classification)*
SEE REMARKS (ON BACK)

12. PERSONAL AUTHOR(S)
Melvin L. Moeschberger and John P. Klein

13a. TYPE OF REPORT
Annual

13b. TIME COVERED
FROM 1/9/82 TO 30/9/83

14. DATE OF REPORT *(Yr., Mo., Day)*
26 OCT 83

15. PAGE COUNT
46

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Modeling series systems; system reliability; competing risks; bivariate exponential distributions; test for independence; consistency of the product limit estimator; (CONTINUED)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The overall objective of this proposal is to investigate the robustness to departures from independence of methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. The first specific aim is to examine the error one makes in modeling a series system by a model which assumes statistically independent component lifetimes when in fact the component lifetimes follow some multivariate distribution. The second specific aim is to assess the effects of the independence assumption on the error in estimating component parameters from life tests on series systems. In both cases, estimates of such errors will be determined via mathematical analysis. A graphical display of the errors for representative distributions will be made available to researchers who wish to assess the possible erroneous assumption of independent competing risks. A third aim is to tighten the bounds on estimates of component reliability when the risks belong to a general dependence class of distributions (for example, (CONTINUED)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. Robert T. Smythe

22b. TELEPHONE NUMBER *(Include Area Code)*
(202) 767-5028

22c. OFFICE SYMBOL
NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

84

ITEM #11, TITLE: EFFECTS OF ASSUMING INDEPENDENT COMPONENT FAILURE TIMES, IF THEY ARE ACTUALLY DEPENDENT, IN A SERIES SYSTEM

ITEM #18, CONTINUED: dependent risks; estimation of component life.

ITEM #19, CONTINUED: positive quadrant dependence, positive regression dependence, etc.). Major decisions involving reliability studies, based on competing risk methodology, have been made in the past and will continue to be made in the future. This study will provide the user of such techniques with a clearer understanding of the robustness of the analyses to departures from independent risks, an assumption commonly made by the methods currently in use.

Annual Report for the
Air Force Office of Scientific Research

by

The Ohio State University Research Foundation
1314 Kinnear Road
Columbus, Ohio 43210

for

Melvin L. Moeschberger
Associate Professor of Preventive Medicine

and

John P. Klein
Assistant Professor of Statistics

Title: Effects of Assuming Independent Component Failure Times, if They are Actually Dependent in a Series System

Period Covered: September 1, 1982 to September 30, 1983

Principal Investigators: Melvin L. Moeschberger and John P. Klein

Phone: (614) 421-3878 or (614) 422-4017

Business Contact: Ralph M. Metzger, III or Charles C. Hall

Phone: (614) 422-2655 or (614) 422-3730

Date of Submission: October 26, 1983

A. Table of Contents

B. TECHNICAL SECTION

I. Abstract

The overall objective of this proposal is to investigate the robustness to departures from independence of methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. The first specific aim is to examine the error one makes in modeling a series system by a model which assumes statistically independent component lifetimes when in fact the component lifetimes follow some multivariate distribution. The second specific aim is to assess the effects of the independence assumption on the error in estimating component parameters from life tests on series systems. In both cases, estimates of such errors will be determined via mathematical analysis and computer simulations for several prominent multivariate distributions. A graphical display of the errors for representative distributions will be made available to researchers who wish to assess the possible erroneous assumption of independent competing risks. A third aim is to tighten the bounds on estimates of component reliability when the risks belong to a general dependence class of distributions (for example, positive quadrant dependence, positive regression dependence, etc.). Major decisions involving reliability studies, based on competing risk methodology, have been made in the past and will continue to be made in the future. This study will provide the user of such techniques with a clearer understanding of the robustness of the analyses to departures from independent risks, an assumption commonly made by the methods currently in use.

II. Specific Objectives

The overall objective is to investigate the robustness to departures from independence to methods currently in use in reliability studies when competing failure modes or competing causes of failure associated with a single mode are present in a series system. We shall also refer to such competitive events as competing risks. The approach will be through the investigation of certain aspects of specific parametric multivariate distributions or by classes of distributions which are appropriate in reliability analyses when there are competing risks present.

The specific objectives are:

1) to assess the error incurred in modeling system life in a series system assumed to have independent component lifetimes when in fact the component lifetimes are dependent.

2) to assess the error in estimating component parameters (i.e., component reliability, mean component life, etc.) in a series system employing either parametric or nonparametric models which assume independent component failure times when in fact the lifetimes are dependent and follow some plausible multivariate distribution.*

3) to derive bounds on component reliability when the failure modes are dependent and fall in a particular dependence class (e.g., positive quadrant dependence, positive regression dependence, etc.).

4) to develop tests of independence, based on data collected from series systems, by making some restrictive assumption about the structure of the systems.**

* A plausible parametric multivariate distribution will be one that satisfies one of the following conditions:

i) the distribution of the minimum of the component failure times closely approximates widely accepted families of system life distributions.

or ii) the marginal distributions closely approximate the distributions of component failure times in the absence of other failure modes.

**This objective has been added to the original objectives because it answers a natural question raised by our preliminary investigation.

III. Introduction to Problem and Significance of Study

Alvin Weinberg (1978) in an editorial comment in the published proceedings of a workshop on Environmental Biological Hazards and Competing Risks noted that "the question of competing risks will not quietly go away: corrections for competing risks should be applied routinely to data." The problem of competing risks commonly arises in a wide range of experimental situations. Although we shall confine our attention in the following discussion to those situations involving series systems in which competing failure modes or competing causes of failure associated with a single mode are present, it is certainly true that we might just as easily speak of clinical trials, animal experiments, or other medical and biological studies where competing events interrupt our study of the main event of interest (cf. Lagokos (1979)).

Consider electronic or mechanical systems, such as satellite transmission equipment, computers, aircraft, missiles and other weaponry consisting of several components in series. Usually each component will have a random life length and the life of the entire system will end with the failure of the shortest lived component. We will examine two situations more closely in which competing risks play a vital role.

First, suppose we are attempting to evaluate system life from knowledge of the individual component lifetimes. Such an evaluation will utilize either an analysis involving mathematical statistics or a computer simulation. At a recent conference on Modeling and Simulation, McLean (1981) presented a scheme to simulate the life of a missile which consisted of many major components in series. The failure distribution associated with each component was assumed to be known (usually exponential or Weibull.) To arrive at the system failure distribution, the components were assumed to act independently of each other. Realistically, this may or may not be the case. If the component lifetimes were dependent for any reason, the computed system failure distribution (as well as its subsequent parameters such as system mean life and system reliability for a specified time) would only crudely approximate the true distribution. The first specific aim of this proposal is to ascertain the error incurred in modeling system life in a series system assumed to have independent component lifetimes (i.e., risks) when, in fact, the risks are dependent.

Second, suppose we wish to evaluate some aspect of the distribution of a particular failure mode based on a typical life test of a series system. The response of interest is the time until failure of a particular mode of interest. Frequently this response will not be observable due to the occurrence of some other event which precludes failure associated with the mode of interest. We shall term such competing events which interrupt our study of the main failure modes of interest as competing risks.

Competing risks arise in such reliability studies when

1) the study is terminated due to a lack of funds or the pre-determined period of observation has expired (Type I censoring).

2) the study is terminated due to a pre-determined number of failures of the particular failure mode of interest being observed (Type II censoring).

3) some systems fail because components other than the one of interest malfunction.

4) the component of interest fails from some cause other than the one of interest.

In all four situations, one may think of the main event of interest as being censored, i.e., not fully observable. In the first two situations, the time to occurrence of the event of interest should be independent of the censoring mechanism. In such instances, the methodology for estimating relevant reliability probabilities has received considerable attention (cf. David and Moeschberger (1978), Kalbfeish and Prentice (1980), Elandt-Johnson and Johnson (1980), Mann, Schafer, Singpurwalla (1974) and Barlow and Proschan (1975) for references and discussion). In the third situation, the time to failure of the component of interest may or may not be independent of the failure times of other components in the system. For example, there may be common environmental factors such as extreme temperature which may affect the lifetime of several components. Thus the question of dependent competing risks is raised. A similar observation may be made with respect to the fourth situation, viz., failure times associated with different failure modes of a single component may be dependent. For a very special type of dependence, the models discussed by Marshall-Olkin (1967), Langberg, Proschan and Quinzy (1978), and Langberg, Proschan, and Quinzy (1981) allow one to convert dependent models into independent ones.

If no assumptions whatever are made about the type of dependence between the distribution of potential failure times, there appears to be little hope of estimating relevant component parameters. In some situations, one may be appreciably misled (cf. Tsiatis (1975), Peterson (1976)). However, as Easterling (1980) so clearly points out in his review of Birnbaum's (1979) monograph

> "there seems to be a need for some robustness studies. How far might one be off, quantitatively, if his analysis is based on incorrect assumptions?"

The second specific aim will address this important issue. First if a specific parametric model which assumes independent risks has been used in the analysis, it would be of interest to know how the error in estimation is affected by this assumption of independence. That is, if independent specific parametric distributions are assumed for the failure times associated with different failure modes when we really should use a bivariate (or multivariate) distribution, then what is the magnitude of the error in estimating component parameters? Secondly, one may wish to allow for a less stringent type of model assumption, and ask the same question with regard to the estimation error. That is, if a nonparametric analysis is performed, assuming independent risks, when some types of dependencies may be present, then what is the magnitude of the estimation error?

The third specific aim will attempt to obtain bounds on the component reliability when the failure times belong to a broad dependence class (e.g., association, positive quadrant dependence, positive regression dependence, etc.). More details will be presented in the methods section.

In summary, competing risk analyses have been performed in the past and will continue to be performed in the future. This study will provide the user of such techniques with a clearer understanding of the robustness to departures from independent risks, an assumption which most of the methods currently in use assume.

IV. Progress Report on First Year's Work

We believe that substantial progress has been made in dealing with the first two objectives as outlined on page 4. We have examined the consequences of erroneously assuming independence when the true multivariate distribution of component lifetimes follows either a Gumbel (1960) bivariate exponential model or the well-known Marshall-Olkin (1967) bivariate exponential model. These bivariate distributions model two series systems which have quite different failure mechanisms. The former has the flexibility of allowing either positive or negative correlation (not due to simultaneous failure) between two component lifetimes in a series system. The latter allows dependency to enter via simultaneous failure of the components. Results associated with both models have been submitted for publication. (See Appendices A and B for copies of these papers.) The paper dealing with the Gumbel model has recently been substantially revised for Technometrics. In addition to clarifying certain aspects of the paper as originally submitted, we have investigated how well the Mann-Grubbs (1974) confidence intervals perform for varying degrees of correlation among the components and for various sample sizes. The Gumbel results were presented in an invited talk at the Applied Statistics Conference in Newark in December, 1982. The Marshall-Olkin results were presented at the joint ASA-IMS-Biometric Society meetings in Toronto in August, 1983. Also, some aspects of these results were given in an invited session on Survival Analysis at the AAAS in Detroit in May, 1983.

We also have examined, in a related issue, the asymptotic bias of the Kaplan-Meier product limit estimator under dependent competing risks and have evaluated the magnitude of this for some specific bivariate distributions. These results have also been submitted for publication. (See Appendices for a copy.)

Work is near completion with respect to evaluating, the consequences of erroneously assuming independence when modeling system reliability from complete component information. This work includes, in addition to the above models, Downton's bivariate exponential, two additional Gumbel models, and the bivariate model proposed by Oakes (1982).

A natural question, which arose from our investigation into the effects of dependence on modeling system reliability and in estimating component parameters from system data, is how can one test, statistically, the assumption of independence? A test of independence, based on data collected from series systems, is not possible, unless some restrictive assumptions are made, due to the nonidentifiability problem. That is, for any set of observable information coming from a series system with dependent component lifetimes there exists a system with independent component lives which has the same observable quantities (c.f. Tsiatis (1975), Rose (1973), Basu and Klein (1982)). By making some restrictive assumptions about the joint distribution of the component lifetimes or about the net component lifetimes it may be possible to test for independence in this restrictive setting.

An approach which makes some assumptions about the marginal component lifetimes and then uses the information to modify the standard nonparametric test for independence is presented in another paper in the Appendix D. The thrust of our future efforts will be to study further aspects of the test system reliabilities under various conditions. In conclusion, we believe that we are on target with respect to dealing with the objectives as outlined in our original proposal as well as the one added objective which serves to enhance the overall effort.

VI. Methods

We refer to pages 8052 of the original proposal for a discussion of the general methodology.

LITERATURE CITED

Ahmed, A.N., Langberg, N.A., Leon, R.V. and Proschan, F. (1978). Two concepts of positive dependence, with applications in multivariate analysis. Statistics Report No. M486, Florida State University, Tallahassee.

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing - Probability Models. Holt, Rinehart and Winston, Inc., New York.

Basu, A.P. and Klein, J.P. (1982). Some recent results in competing risks. Survival Analysis. Crowley and Johnson, Editors, 216-229.

Berman, S.N. (1963). Notes on extreme values, competing risks, and semi-Markov processes. Ann. Math. Statist. 34, 1104-1106.

Birnbaum, Z.W. (1979). On the Mathematics of Competing Risks. U.S. Dept. of HEW. Publication No. 79-1351.

Chiang, C.L. (1968). Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

Cox, D.R. (1962). Renewal Theory. Methuen, London.

David, H.A. (1974). Parametric approaches to the theory of competing risks. Pp. 275-290 in: Proschan and Serfling (1974).

David, H.A. and Moeschberger, M.L. (1978). Theory of Competing Risks. Griffin, London.

Easterling, R.G. (1980). Book review of Z.W. Birnbaum monograph. Technometrics 22: 131-132.

Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival Models and Data Analysis. Wiley, New York.

Fisher, L. and Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. Reliability and Biometry, Statistical Analysis of Lifelength. Siam, 303-326.

Fishman, G.S. (1973). Concepts and Methods in Discrete Event Digital Simulation. Wiley, New York.

Friday, D.S. and Patel, G.P. (1977). A bivariate exponential model with applications to reliability and computer generation of random variates. Theory and Applications of Reliability, Vol. I. C.P. Tsokos and I.N. Shimi, Editors, 527-548.

Gail, M. (1975). A review and critique of some models used in competing risk analysis. Biometrics 31, 209-222.

Gibbons, J.D. (1971). Nonparametric Statistical Inference. McGraw-Hill Book Company, New York.

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55: 698-707.

Hoel, D.G. (1972). A representation of mortality data by competing risks. Biometrics 28, 475-488.

Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Klein, J.P. and Basu A.P. (1981). Replacing Dependent Systems by Independent Systems in a Competing Risk Framework with Applications'. Submitted for publication.

Kotz, S. (1974). Multivariate Distributions at a Cross Road in Statistical Distributions in Scientific Work VI. Reidell Publishing Co., Boston, 247-270.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35: 139-156.

Langberg, N., Proschan, F. and Quinzi, A.J. (1978). Converting Dependent Models into Independent Ones, Preserving Essential Features Ann Prob., 6, 174-181.

Langberg, N., Proschan, F. and Quinzi, A.J. (1981). Estimating Dependent Lifelengths with Applications to the Theory of Competing Risks . Ann. Statist. 9:157-167.

Lee L. and Thompson, W.A., Jr. (1974). Results on failure time and pattern for the series system. Pp 291-302 in: Proschan and Serfling (1974).

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics 16:335-347.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974) Methods for the Statistical Analysis of Reliability and Life Data, Wiley, N.Y.

Marshall, A.W. and Olkin, I. (1967). A Multivariate Exponential Distribution J. Amer. Statist. Assoc. 66, 30-40.

McLean, T.J. (1981). Customer's Risk Evaluation. Paper presented to Modeling and Simulation Conference. Pittsburgh, Pa.

Mendenhall, W. and Lehman, E.H., Jr. (1960). An approximation to the negative moments of the positive binomial useful in life testing. Technometrics 2:227-242.

Miller, D.R. (1977). A note on independence of multivariate lifetimes in competing risks. Ann Statist.5: 576-579.

Moeschberger, M.L. and David, H.A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-933.

Moeschberger, M.L. (1974). Life tests under dependent competing causes of failure. Technometrics 16, 39-47.

Oakes, D. (1982). A model for association in bivariate survival data. J. Roy. Statis. Soc. 44:414-422.

Peterson, A.V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73: 11-13.

Peterson, A.V., Jr. (1978). Dependent competing risks: bounds for net survival functions with fixed crude survival functions. Environmental International 1: 351-355.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Jr., Flournoy, N., Farewell, V.T., and Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34: 541-554.

Proschan, F. and Serfling, R.J. (Eds.) (1974). Reliability and Biometry: Statistical Analysis of Lifelength. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Rose, D.M. (1973). Investigation of dependent competing risks. Ph.D. dissertation, University of Washington.

Thompson, W.A., Jr. (1979). Technical note: competing risk presentation of reactor safety studies. Nuclear Safety 20: 414-417.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. 72, 20-22.

Tsiatis, A. (1977). Nonidentifiability problems with the reliability approach to competing risks. Technical Report No. 490, University of Wisconsin-Madison.

Weinberg, Alvin (1978). Editorial. Environmental International 1:285-287.

APPENDIX A

CONSEQUENCES OF DEPARTURES FROM INDEPENDENCE IN EXPONENTIAL SERIES SYSTEMS

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

30 September, 1983

Annual Report for Period 1 September 1982 - 30 September, 1983

AFOSR - NM
Building 410
Bolling Air Force Base
Washington DC 20332

AFOSR-PKZ
Building 410
Bolling Air Force Base
Washington DC 20332

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Consequences of Departures from Independence in Exponential Series Systems		5. TYPE OF REPORT & PERIOD COVERED Annual Sept. 1 1982 to Sept.30 1983
		6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763265
7. AUTHOR(s) M. L. Moeschberger and John P. Klein		8. CONTRACT OR GRANT NUMBER(s) AFSOR- 82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington DC 20332		12. REPORT DATE 30 Sept., 1983
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington DC 20332		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited; approved for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

competing risks, component life, modeling series systems, robustness studies, system reliability, Gumbel bivariate exponential

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reports the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow a Gumbel multivariate exponential in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life is studied based on lifetest data collected on series systems when the assumption of independence is

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

made erroneously.

Consequences of Departures from Independence in Exponential Series Systems

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

Abstract

This paper investigates the consequences of departures from independence when the component lifetimes in a series system are exponentially distributed. Such departures are studied when the joint distribution is assumed to follow a Gumbel bivariate exponential model. Two distinct situations are considered. First, in theoretical modeling of series systems, when the distribution of the component lifetimes is assumed, one wishes to compute system reliability and mean system life. Second, errors in parametric and nonparametric estimation of component reliability and component mean life are studied based on lifetest data collected on series systems when the assumption of independence is made erroneously.

Keywords: competing risks, component life, modeling series systems, robustness studies, system reliability, Gumbel bivariate exponential

1. INTRODUCTION

Consider a system consisting of several components linked in series. For such a system the failure of any one of the components causes the system to fail. Common assumptions made in modeling and analyzing data from such a system are that the component lifetimes are independent and exponentially distributed. Many authors have considered the problem of analyzing a series system with exponential component lives. For example, confidence bounds for system reliability assuming independent exponentially distributed component lifetimes were presented in Mann(1974) and Mann and Grubbs (1974). (cf., see Mann, Schafer, and Singpurwalla (1974) for a more comprehensive review.) More recently, work invoking the assumption of independent exponentially distributed lifetimes has been presented by Chao (1981) and Miyamura (1982). Estimation of component parameters from series system data has been treated by Boardman and Kendell (1970) in the context of independent exponential component lives. Some authors suggest a nonparametric alternative to the estimation of component reliability based on series system data (cf. Kalbfleisch and Prentice (1980) and Lawless (1982)).

The assumption of independence is essential to these analyses and also an important concern. Several authors have shown that this assumption, by itself, is not testable since, based on data from a series system, there is no way to distinquish between an independent and dependent model. (See Tsiatis (1975), Peterson (1976), and Basu (1981) for a discussion of nonidentifiability results.) In many situations one may be appreciably misled by the independence assumption.

Lagakos (1979), in a study of the effects of various types of dependence among component lifetimes, has noted that most methods of analysis have assumed noninformative models of which independence is a special case. He points out, "it is important to be aware of the possible consequences of making this assumption when it is false." Furthermore, Easterling (1980) states in his review of Birnbaum's (1979) monograph on competing risks "there seems to be a need for some robustness studies. How far might one be off, quantitatively, if his analysis is based on incorrect assumptions?"

In this paper we consider the consequences of departures from independence when the component lifetimes are exponentially distributed. Such departures from independence may be related to some common environmental factor, only present when the components are linked together in series. The load each component is subject to is either reduced or increased according to the age of the system. To study such departures we have selected a model proposed by Gumbel (1960). Gumbel's model retains the assumption of exponentially distributed component lifetimes while allowing the flexibility of both positive or negative mild correlation between component lifetimes.

The effects of a departure from the assumption of independent component lifetimes in a series system will be addressed in two distinct situations. The first situation arises in modeling the performance of a theoretical series system constructed from components whose lifetimes are exponentially distributed. Here, based on testing each component separately or based on engineering design principles, it is reasonable to assume that the components are known to be exponentially distributed with known parameter values. Based on this information we wish to calculate parameters such as the mean system life or system reliability of a series system constructed from these components. In Section 2 we describe how the values of these quantities are affected by departures from independence when the component parameters are completely specified. In Section 3 we study the performance of the Mann-Grubbs (1974) confidence bounds on system reliability for small sample sizes and for varying degrees of correlation when the component parameters are estimated from component data.

The second situation involves making inferences about component lifetimes distributions, component reliabilities, and component mean lives, from data collected on series systems. Commonly, data collected on such systems are analyzed by assuming a constant-sum model of which independence is a special case (cf. Williams and Lagakos (1977) and Lagakos and Williams (1978). In Section 4 we study the properties of the maximum likelihood estimators of component parameters calculated under an assumption of independent exponential component lifetimes when the component lifetimes are Gumbel bivariate exponential. Because of the usage of nonparametric estimates of component reliability, we also study the estimation error of the Kaplan-Meier (1958) estimator when the assumption of independence is made erroneously.

2. MODELING SYSTEM RELIABILITY FROM COMPLETE COMPONENT INFORMATION

Consider a two component series system with component lifelengths X_1, X_2. Suppose that X_i has an exponential survival function

$$\bar{F}_i(t) = P(X_i > t) = \exp(-\lambda_i t),$$
$$\lambda_i,\ t>0,\ i = 1,2.$$

This assumption is made on the basis of extensive testing of each component separately or on knowledge of the underlying mechanism of failure. The value of λ_i is assumed known. If (X_1,X_2) are independent, then the time to system failure has an exponential distribution with failure rate $\lambda=\lambda_1+\lambda_2$, and the system reliability is given by

$$\bar{F}_I(t) = P[\min(X_1,X_2)>t \mid \text{independence}] = \exp(-\lambda t). \quad (2.1)$$

Suppose that the actual joint distribution of (X_1,X_2) has the form proposed by Gumbel (1960), namely,

$$P(X_1>x_1, X_2>x_2)$$
$$=[\exp(-\lambda_1 x_1-\lambda_2 x_2)]\,[1+\alpha\,(1-\exp(-\lambda_1 x_1))\,(1-\exp(-\lambda_2 x_2))] \quad (2.2)$$

and the joint probability density of (X_1,X_2) is

$$f(x_1,x_2)=\lambda_1\lambda_2[\exp(-\lambda_1 x_1-\lambda_2 x_2)]\,[1+\alpha(2\exp(-\lambda_1 x_1)-1)(2\exp(-\lambda_2 x_2)-1)] \quad (2.3)$$

where in both (2.2) and (2.3), $x_1,x_2,\lambda_1,\ \lambda_2>0, -1\le\alpha\le 1$. This distribution has marginal survival functions equivalent to those for the independent model which, in part, is the reason for choosing it. The correlation between (X_1,X_2) is $\rho=\alpha/4$ and $\alpha=0$ is equivalent to (X_1,X_2) being independent. For $\rho>0$ (<0) the components are positively (negatively) quadrant dependent (cf. Barlow and Proschan (1975)). Furthermore, the conditional expectation of X_1, given $X_2=x_2$, is

$$E(X_1|X_2=x_2)=\frac{1}{\lambda_1}\left[1+2\rho-4\rho\exp(-\lambda_2 x_2)\right].$$

If (X_1,X_2) have the joint distribution (2.3) then the true system reliability, is

$$\bar{F}_D(t) = P\left[\min(X_1,X_2)>t\,|\,\text{dependence}\right]$$
$$= \exp(-\lambda t)\left[1+4\rho(1-\exp(-\lambda_1 t))\ (1-\exp(-\lambda_2 t))\right]. \qquad (2.4)$$

From (2.1) and (2.4) we can easily see the amount of error in modeling system reliability is

$$\Delta(t)=\bar{F}_D(t)-\bar{F}_I(t)$$
$$=4\rho\left[1-\exp(-\lambda_1 t)\right]\left[1-\exp(-\lambda_2 t)\right]\ \exp\left[-(\lambda_1+\lambda_2)t\right]. \qquad (2.5)$$

Note that $|\Delta(t)|$ increases as $|\rho|$ increases, for fixed λ_1,λ_2, and t. The magnitude of $\Delta(t)$, of course, depends upon λ_1,λ_2, t, and ρ. When $\lambda_1=\lambda_2=\phi$, one can show that $\Delta(t)$ is maximized at $t=\left[\ln 2\right]/\phi$ (fixing ρ and ϕ). The value of $|\Delta(t)|$ at this point is $|\rho|/4$ which is at most 1/16. Representative values of $\bar{F}_D(t)$ for $\lambda_1=1,\lambda_2=1.5$ and $\rho=-.25,-.125,0,\ .125,\ .25$ are plotted in Figure 1. The curve with $\rho=0$ corresponds to the system reliability if the assumption of independence is true. Since most applications of interest involve reliabilities of .75 or greater, we plot in Figure 2 the ratio of the 100 p^{th} upper percentiles under dependence and independence vs. the correlation. From Figure 2, it appears that,

Figure 1. System Reliability for Gumbel's Model
$\lambda_1 = 1., \lambda_2 = 1.5.$

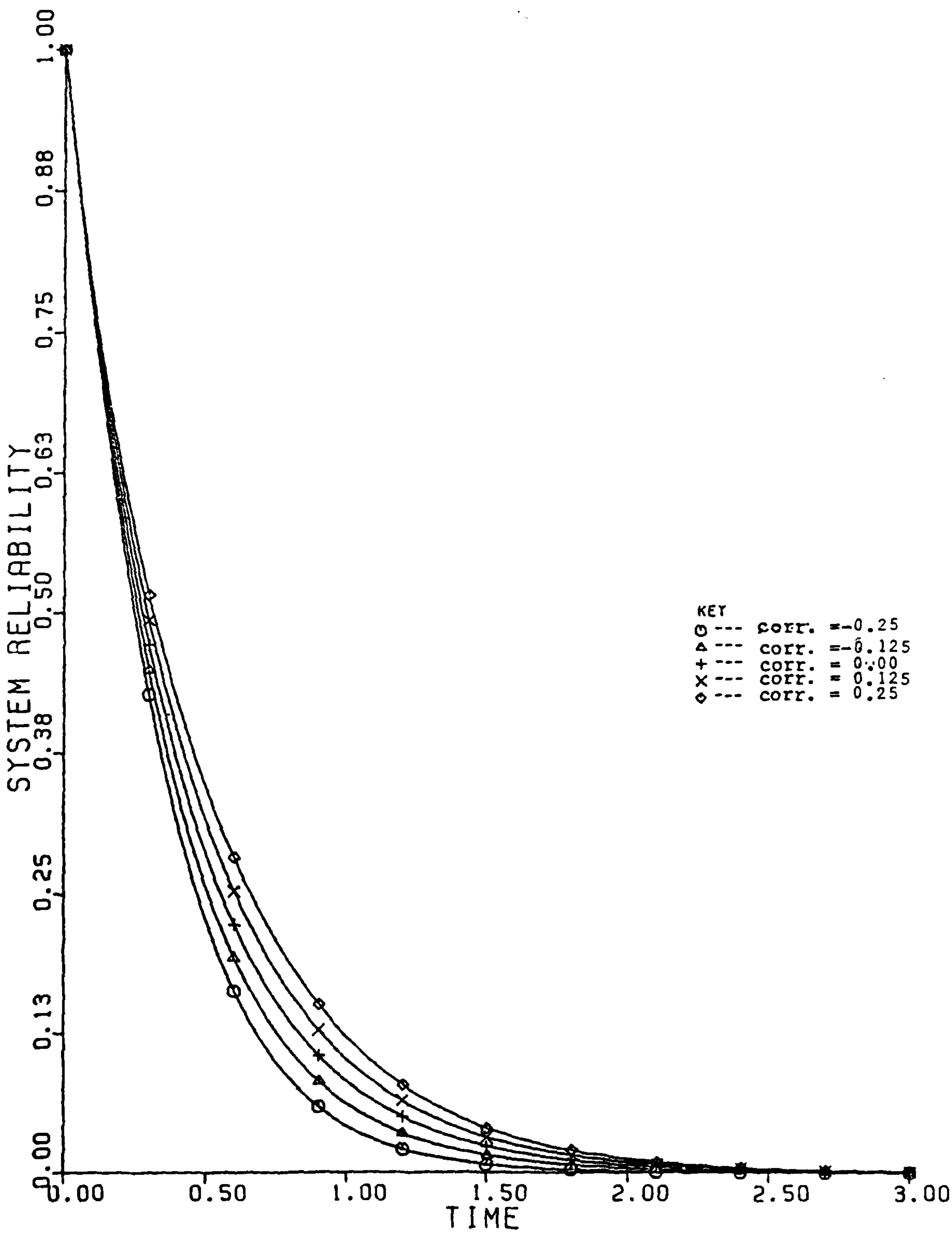
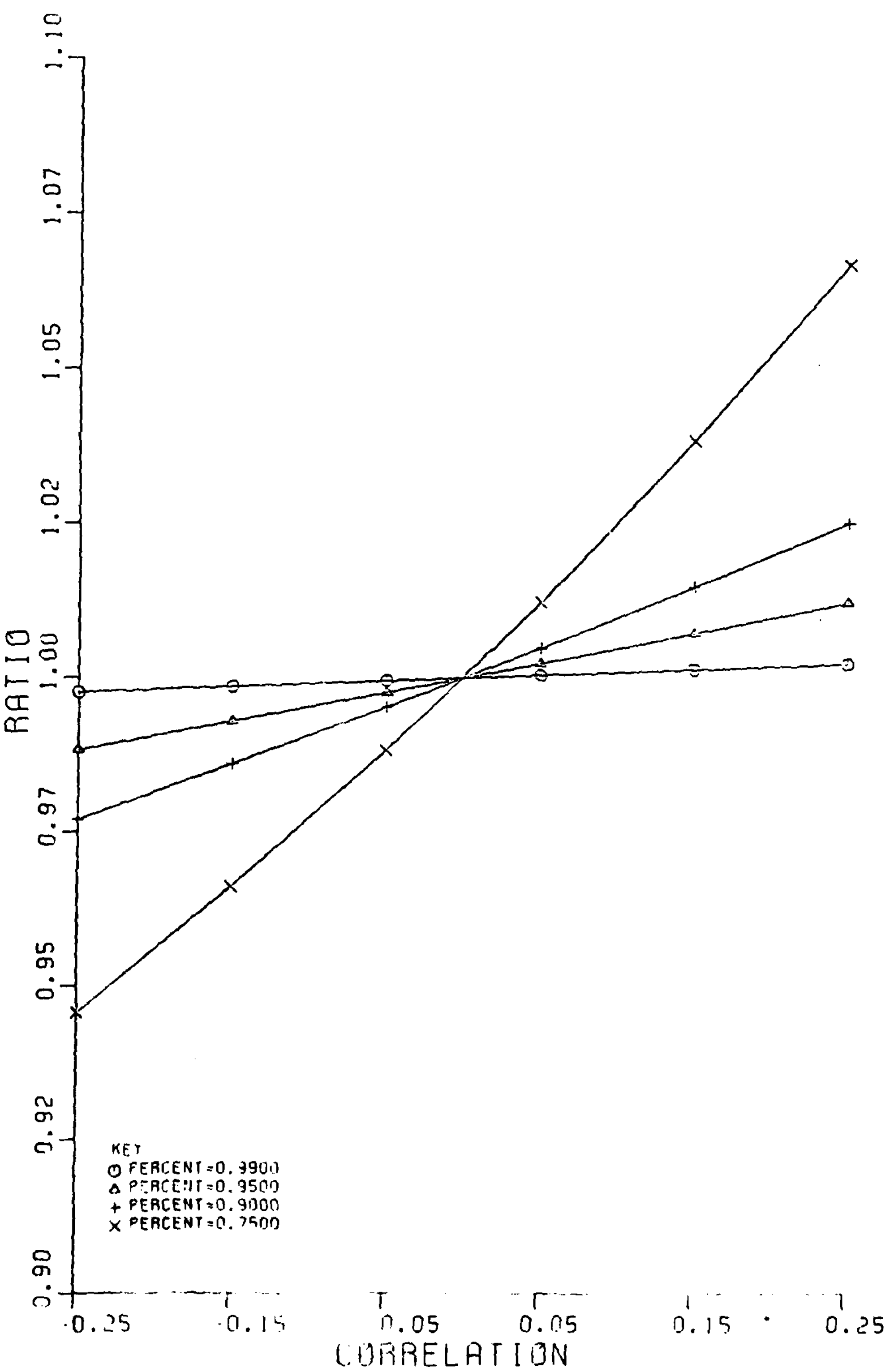


Figure 2. Ratio of $100p^{th}$ percentile under dependence and independence vs. correlation for $\lambda_1=1.0$, $\lambda_2=1.5$.



when the predicted system reliability under independence is greater than .90, misspecifying the dependence parameter has little effect. However, in the range where the predicted system reliability under independence is less than .75, misspecifying the dependence parameter may lead to errors exceeding 6%. Maximum values of $|\Delta(t)|$ are tabled in Table 1, for $\lambda_1=1$ and various values of λ_2.

The mean time to system failure based on (2.1), assuming independence, is

$$\mu_I = 1/(\lambda_1+\lambda_2) \qquad (2.6)$$

and that based on (2.4) is

$$\mu_D = \frac{1}{(\lambda_1+\lambda_2)} + 4\rho\left[\frac{3}{2(\lambda_1+\lambda_2)} - \frac{1}{(2\lambda_1+\lambda_2} - \frac{1}{(\lambda_1+2\lambda_2)}\right] \qquad (2.7)$$

The amount of error in modeling system mean life is

$$\mu_D-\mu_I = \frac{6\rho\lambda_1\lambda_2}{(\lambda_1+\lambda_2)\ (2\lambda_1+\lambda_2)\ (\lambda_1+2\lambda_2)}$$

$$= \frac{6\rho\lambda_1\lambda_2\mu_I}{(2\lambda_1+\lambda_2)\ (\lambda_1+2\lambda_2)} \qquad (2.8)$$

whose absolute value obviously increases as $|\rho|$ increases. If $\lambda_1=\lambda_2$, this error reduces to $2\rho\mu_I/3$ which has a maximum absolute value of $\mu_I/6$.

It is apparent from Table 1 and equations (2.5) and (2.8) that the error in modeling system reliability and mean system life, based on independence, increases as $|\rho|$ increases and is a function of the relative sizes of λ_1 and λ_2. In particular, when the mean life of one component is substantially greater than the mean life of the second component, then the behavior of the system is well-approximated by the behavior of the shorter-lived component acting alone. This can be seen in (2.4) and (2.7) by letting $\lambda_1 \to 0$. In this instance, we see also, from (2.5) and (2.8), that the amount of error incurred by assuming independence is negligible.

3. ESTIMATING SYSTEM RELIABILITY FROM COMPONENT DATA

A common practice in predicting system reliability is to test each of the components independently and then to use the data to obtain confidence bounds on system reliability. These bounds, obtained by Mann and Grubbs (1974), assume that the component lifetimes are exponential, and that the components act independently when linked in series. In the bivariate case the bounds are computed as follows: for the j^{th} component suppose that n_j prototypes have been tested until $(r_j \leq n_j)$ failures occur.

Let Z_j be the total time on test for the jth component. Define

$$M^* = \sum (r_j - 1)/Z_j + \left\{ \sum (r_j - 1/Z_j^3 \right\} / \left\{ \sum (r_j - 1)/Z_j^2 \right\}, \qquad (3.1)$$

and

$$V^* = \sum (r_j - 1)/Z_j^2 + \left\{ \sum (r_j - 1)/Z_j^4 \right\} / \left\{ \sum (r_j - 1)/Z_j^2 \right\}. \qquad (3.2)$$

An approximate γ level lower confidence level for system reliability at time t_m is

$$\exp\left[-t_m M^* \left\{ 1 - V^*/(9M^{*2}) + \eta_\gamma (V^*)^{\frac{1}{2}}/(3M^*) \right\}^3 \right], \qquad (3.3)$$

where η_γ is the 100γ percentile of a standard normal random variable.

Table 1. Maximum values of $|\Delta(t)|$ for $\lambda_1=1$ and various values of λ_2.

λ_2	Max $|\Delta(t)|$
2	.056
4	.041
8	.025
16	.014

When the system being evaluated has dependent components these bounds may be misleading. The problem is that component data is independent, since the components are tested separately, but when put together into a system some interdependence may develop. Of course such dependence is not detectable, in the absence of some system data, since the data on components we see are independent. To study the performance of the bound (3.3) when the correct system model is the Gumbel model (2.2) a simulation study was performed. For each simulated sample n_j observations from exponential populations with mean $1/\lambda_j$, j=1,2, were simulated. The two samples were generated independently. The confidence bound (3.3) was obtained. This was then compared to the true system reliability at various ρ's obtained from (2.4). Ten thousand such bounds were simulated for each set of parameter values. The results for $n_1=n_2=3,5,10, \lambda_1=1.0, \lambda_2=1.5$, at $t_m=0.1$ are reported in Table 2. Here the system reliability ranges from .7684 at $\rho=-.25$ to .7891 at $\rho=.25$ with a value of .7788 at $\rho=0$.

The results in Table 2 show that at high negative correlations the coverage probabilities are significantly lower than claimed under independence, while for a high positive correlation the intervals are conservative. This trend becomes more exaggerated as n_1, n_2 increase. This is due to the fact that as n_1, n_2 increases the bound approaches the reliability under independence. As seen in Section 1, the true reliability at t is an increasing function of ρ so that asymptotically coverage probabilities approach 0(or1) for $\rho<0(>0)$. On the practical side, while the

Table 2. Estimated Coverage Probabilities for Mann-Grubbs Bounds
$\lambda_1=1.0$, $\lambda_2=1.5$

n_1	n_2	γ	Correlation -.25	-.15	-.05	0	.05	.15	.25
3	3	.95	93.41-	94.11-	94.74	95.05	95.27	95.80+	96.22+
3	3	.90	87.40-	88.42-	89.32-	89.78	90.20	91.18+	92.15+
3	3	.75	71.03-	72.53-	74.13-	74.88	76.58	77.34+	78.81+
5	5	.95	93.19-	94.04-	94.90	95.26	95.62+	96.17+	96.81+
5	5	.90	87.12-	88.48-	89.85	90.39	91.10	92.13+	93.14
5	5	.75	69.68-	72.02-	74.10-	75.13	76.14+	78.32+	80.03
10	10	.95	92.03-	93.42-	94.58	95.08	95.51+	96.42+	97.14
10	10	.90	85.93-	87.70-	89.34-	90.21	91.05+	92.56+	93.93
10	10	.75	67.77-	70.90-	74.12-	75.63	77.05+	79.87+	82.56

\+ At least two standard errors above specified level

\- At least two standard errors below specified level

Note: Standard errors of above estimates are approximately:
.2 for .95 level, .3 for .90 level, .4 for .75 level

estimated coverage probabilities for $\rho<0$ are statistically significantly lower than expected for small sample sizes, they are not of sufficient magnitude to cause great concern, especially at $\gamma=.95$.

4. ESTIMATING COMPONENT PARAMETERS

In this section we are interested in examining how the independence assumption affects the magnitude of the estimation error in estimating component reliability and mean life from data collected on series systems. That is, for each system put on test we observe its failure time and an indicator variable which tells us which component caused the system to fail. We are interested in how varying degrees of dependence affect the bias and mean squared error of either parametric or nonparametric estimators of component survival and mean life which were obtained by assuming independent component lifetimes.

4.1 Parametric Approach

Consider, first, the problem of parameter estimation in a two component series system. We assume the two components' survival functions are $\bar{F}_i(t) = \exp(-\lambda_i t)$, i = 1,2, and a life test is conducted by putting n systems on test. We observe n_i systems to fail due to failure of the ith, i = 1,2, component. Let T denote the sum of all n failure times. From Moeschberger and David (1971) the maximum likelihood estimator of λ_i is, assuming independence,

$$\hat{\lambda}_i = n_i/T, \ i = 1,2$$

so the estimator of component mean life, $\mu_i=\lambda_i^{-1}$, is

$$\hat{\mu}_i = T/n_i \quad \text{if} \quad n_i>0. \qquad (4.1)$$

Now suppose that we are in fact sampling from the Gumbel distribution (2.3). For this model, component mean life is the same as in the independent case. The random variables (n_i, T) are independent (the conditional distribution of T given n_i is free of n_i) and n_i is binominal with parameters n and $p_i = P(\min(X_1, X_2) = X_i)$. For this model

$$p_1 = P(X_1 < X_2) = \lambda_1 \{ \frac{1}{\lambda_1+\lambda_2} + \frac{4\rho(\lambda_1-\lambda_2)\lambda_2}{(\lambda_1+\lambda_2)(2\lambda_1+\lambda_2)(\lambda_1+2\lambda_2)} \} \quad (4.2)$$

with $p_2 = 1-p_1$. From Mendenhall and Lehman (1960) approximations to the moments of $1/n_i$, conditional on $n_i > 0$, are

$$E(1/n_i | n_i > 0) = \frac{n-2}{n(a-1)}, \quad (4.3)$$

$$E(1/n_i^2 | n_i > 0) = \frac{(n-2)(n-3)}{n^2(a-1)(a-2)} \quad (4.4)$$

where $a = (n-1)p_i$.

The expected value of T is given by $n\mu_D$ where μ_D is given by (2.7) and

$$E(T^2) = n\left[\frac{2+10\rho}{(\lambda_1+\lambda_2)^2} - 8\rho\left(\frac{1}{(2\lambda_1+\lambda_2)^2} + \frac{1}{(\lambda_1+2\lambda_2)^2}\right)\right] + n(n-1)\mu_D^2 . \quad (4.5)$$

Thus the bias and MSE of $\hat{\mu}_i$, conditional on $n_i > 0$, under this model are

$$B(\hat{\mu}_i) = E(\hat{\mu}_i - \mu_i) = \frac{(n-2)\mu_D}{[(n-1)p_i - 1]} - \mu_i, \quad (4.6)$$

and $$MSE(\hat{\mu}_i) = E(T^2)\, E(1/n_i^2 | n_i > 0) - \frac{2\mu_i(n-2)\mu_D}{[(n-1)p_i - 1]} + \mu_i^2, \quad (4.7)$$

for i = 1,2.

We note that for large samples

$$\lim_{n \to \infty} B(\hat{\mu}_i) = \frac{\mu_D}{p_i} - \mu_i \qquad (4.8)$$

$$\lim_{n \to \infty} MSE(\hat{\mu}_i) = (\lim_{n \to \infty} B(\hat{\mu}_i))^2 \qquad (4.9)$$

for i = 1,2.

For $\lambda_1=\lambda_2$ from (4.6), we see that

$$B(\hat{\mu}_i)= \frac{1+2(n-2)\rho/3}{(n-3)} \mu_1 = \frac{\mu_1}{n-3} + \frac{2(n-2)\rho\mu_1}{3(n-3)} . \qquad (4.10)$$

A similar expression holds for $B(\hat{\mu}_2)$. Note that (4.10) consists of two terms. The first term, reflecting sampling error, is positive for all n and dominates the bias expression for small n. The second term, reflecting modeling error, takes on the same sign as the correlation and dominates for large n, approaching the limit of $2\rho\mu_1/3$.

For $\lambda_1 \neq \lambda_2$, the bias can again be expressed as the sum of a term reflecting sampling error and a term reflecting modeling error. The sampling error depends upon the expected number of component 1 failures, np_1. From (4.2) we note that p_1 increases as increases if $\lambda_1>\lambda_2$ and decreases if $\lambda_1<\lambda_2$. The modeling error is reflected in the asymptotic bias which is an increasing function of ρ. A typical representation of the bias of $\hat{\mu}_1$, as a function of sample size and ρ, is in Figure 3 where $\lambda_1=1$, $\lambda_2=1.5$.

Comments similar to those pertaining to the bias may be made for the mean squared error of $\hat{\mu}_1$. Figure 4 depicts the ratio of the MSE $(\hat{\mu}_1|\rho)$ and MSE $(\hat{\mu}_1|\rho=0)$ as a function of sample size and for $\lambda_1=1,\lambda_2=1.5$.

Figure 3. Bias of $\hat{\mu}_1$ Under Gumbel's Model for λ_1=1. and λ_2=1.5

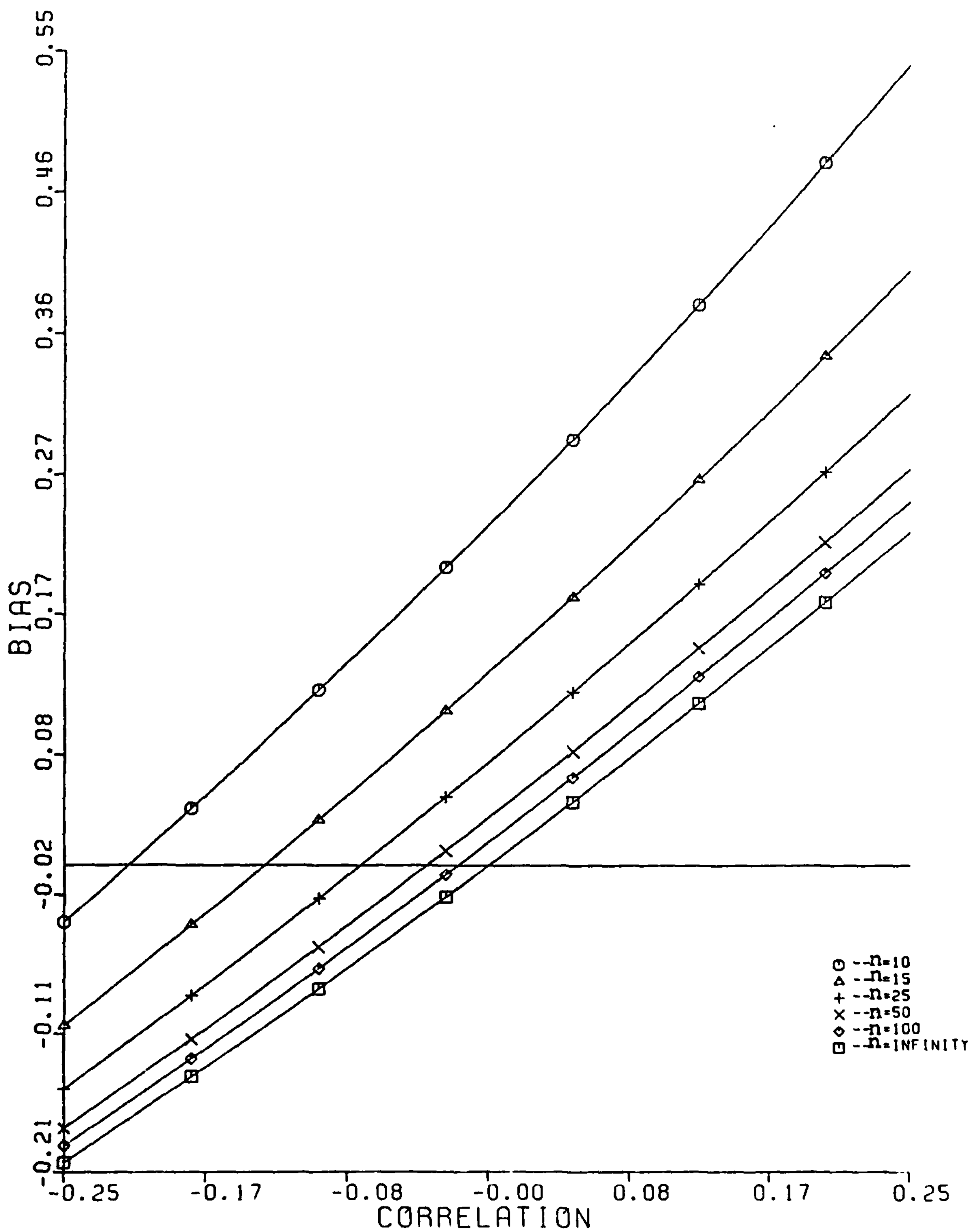
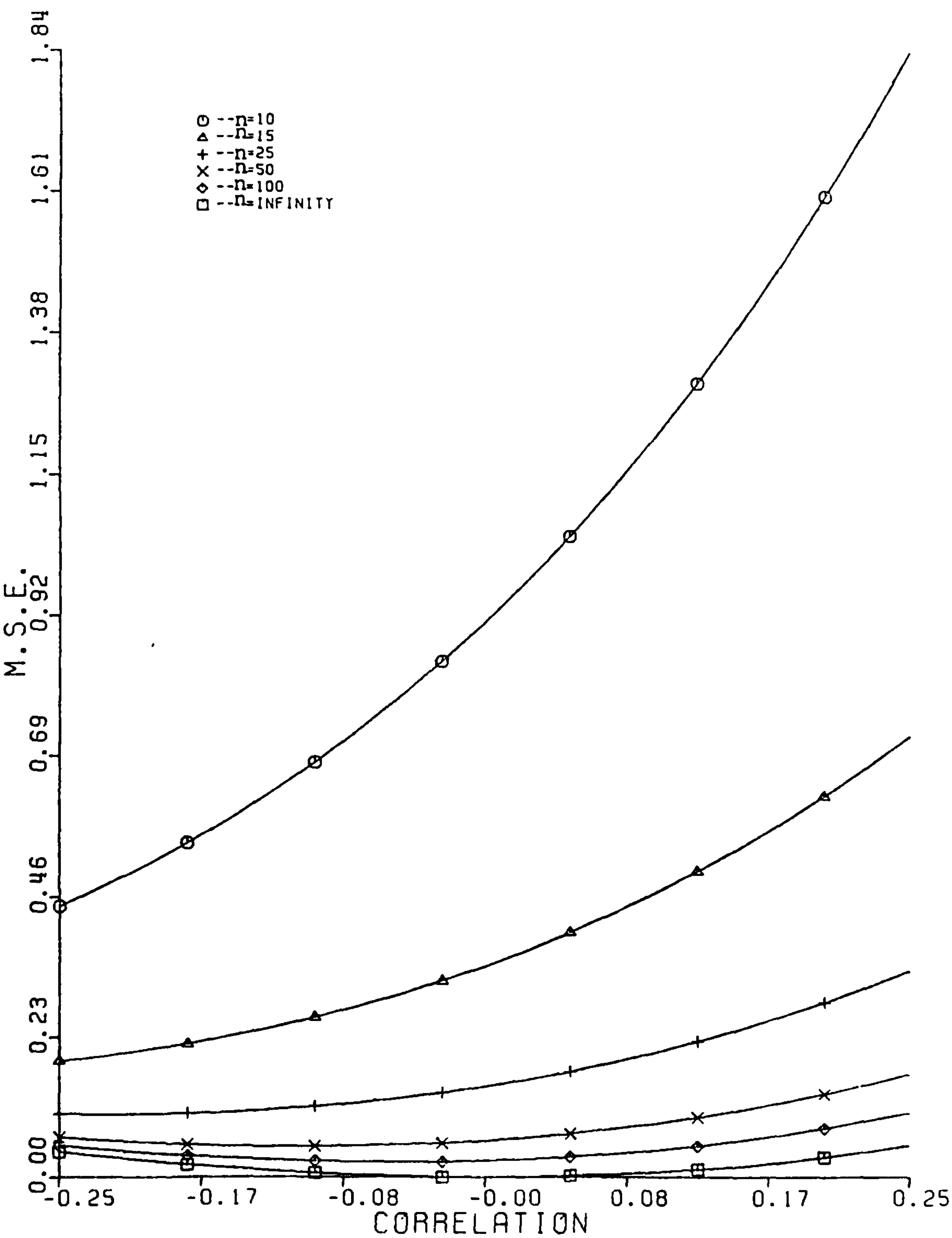


Figure 4. Mean Squared Error of $\hat{\mu}_1$, for Gumbel's Model λ_1=1. and λ_2=1.5.



4.2 Nonparametric Approach

A second approach to the problem of estimating component parameters is via the nonparametric estimator proposed by Kaplan and Meier (1958). The product limit estimator, assuming independent risks, is constructed as follows. As before suppose n systems are put on test at time 0 and n_i systems fail due to failure of component i. Let $X_{i(1)},\ldots,X_{i(n_i)}$ denote the ordered times at which these n_i events occur and let $r_{i1}, \ldots, r_{in_i}$ be the ranks of those ordered survival times among all n ordered lifetimes. The component reliability for the ith component at time x may now be estimated by the product of the individual conditional survival probabilities, namely by

$$\tilde{\bar{F}}_i(x) = \begin{cases} 1 \text{ if } x<x_{i(1)} \\ \prod_{j=1}^{j(i,x)} (n-r_{ij})/(n-r_{ij}+1) & x>x_{i(1)}, \end{cases}$$

where j(i, x) is the largest value of j for which $x_{i(j)}<x$. A special note is needed to cover the case of $x_{i(n_i)}$ not being the largest observed death or removal time. To avoid this problem we shall define $\tilde{\bar{F}}_i(x) = 0$ for x greater than the largest observed failure time.

If the component lifetimes, in fact, follow the Gumbel bivariate exponential, Klein and Moeschberger (1983) have shown that the Kaplan-Meier estimator is not consistent. For i = 1, the Kaplan-Meier estimator is not estimating $\bar{F}_1(t)$ but rather another survival function, $\bar{H}_1(t)$ given by

$$\bar{H}_1(t) = \exp\{-\lambda_1 \int_0^t \frac{[1+4\rho(1-e^{-\lambda_2 u})(1-2e^{-\lambda_1 u})]}{[1+4\rho(1-e^{-\lambda_1 u})(1-e^{-\lambda_2 u})]}\,du\}, \quad t>0 \qquad (4.11)$$

Note that equation (4.11) simplifies, if $\lambda_1=\lambda_2=\phi$, to

$$\bar{H}_1(t) = e^{-\phi t}\left[1+4\rho(1-e^{-\phi t})^2\right]^{1/2} \qquad (4.12)$$

which is increasing in ρ . Similarly, $\tilde{\bar{F}}_2(t)$ is actually estimating $\bar{H}_2(t)$, which is defined analogously.

Measures of the error in estimating $\bar{F}_i(t)$ by $\tilde{\bar{F}}_i(t)$ are again the bias and mean squared error of $\tilde{\bar{F}}_i(t)$ computed under the dependence model. Under this model the Kaplan-Meier estimator is equivalent to the estimator one would obtain based on n observations from an independent system with component survival distributions $\bar{H}_i$ given by (4.11) or, if $\lambda_1=\lambda_2$, by (4.12). Hence from Kaplan and Meier (1958) it follows that the variance of $\tilde{\bar{F}}_i(t)$ is given by

$$V(\tilde{\bar{F}}_i(t)) = \bar{H}_i(t)^2 \int_0^t \frac{|d\bar{H}_i(u)|}{n\,\bar{H}_i(u)^2} \; . \qquad (4.13)$$

Thus the bias and MSE of $\tilde{\bar{F}}_i(t)$ are, from (4.11) and 4.13)

$$B(\tilde{\bar{F}}_i(t))=\bar{H}_i(t) - \bar{F}_i(t),\; t \geq 0, \qquad (4.14)$$

and $$MSE(\tilde{\bar{F}}_i(t)) = (\bar{H}_i(t)-\bar{F}_i(t))^2 + \bar{H}_i(t)^2 \int_0^t \frac{|d\bar{H}_i(u)|}{n\,\bar{H}_i(u)^2}, t>0. \qquad (4.15)$$

The estimator is not consistent since $B(\tilde{\bar{F}}_i(t))$ is independent of n, and not necessarily zero. Also $MSE(\tilde{\bar{F}}_i(t))$ consists of a factor which depends only on the model error and is free of sample size, and, a term which tends to 0 as n tends to infinity.

Figure 5. Bias of Kaplan-Meier Estimate, $\tilde{\tilde{F}}_1(t)$, λ_1=1., λ_2=1.5

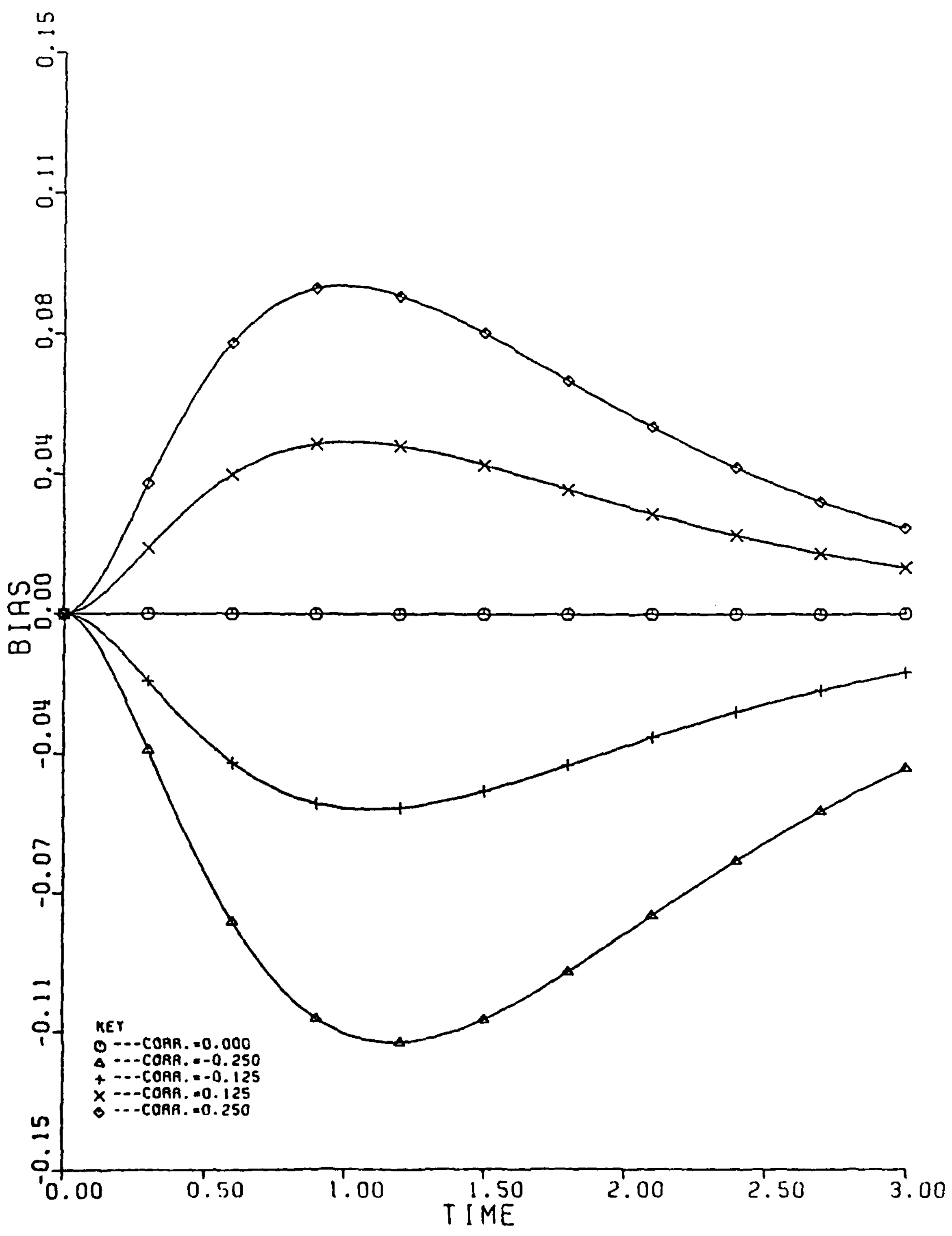


Figure 6A. MSE of Kaplan-Meier Estimate $\tilde{F}_1(t)$, λ_1=1., λ_2=1.5, n=10

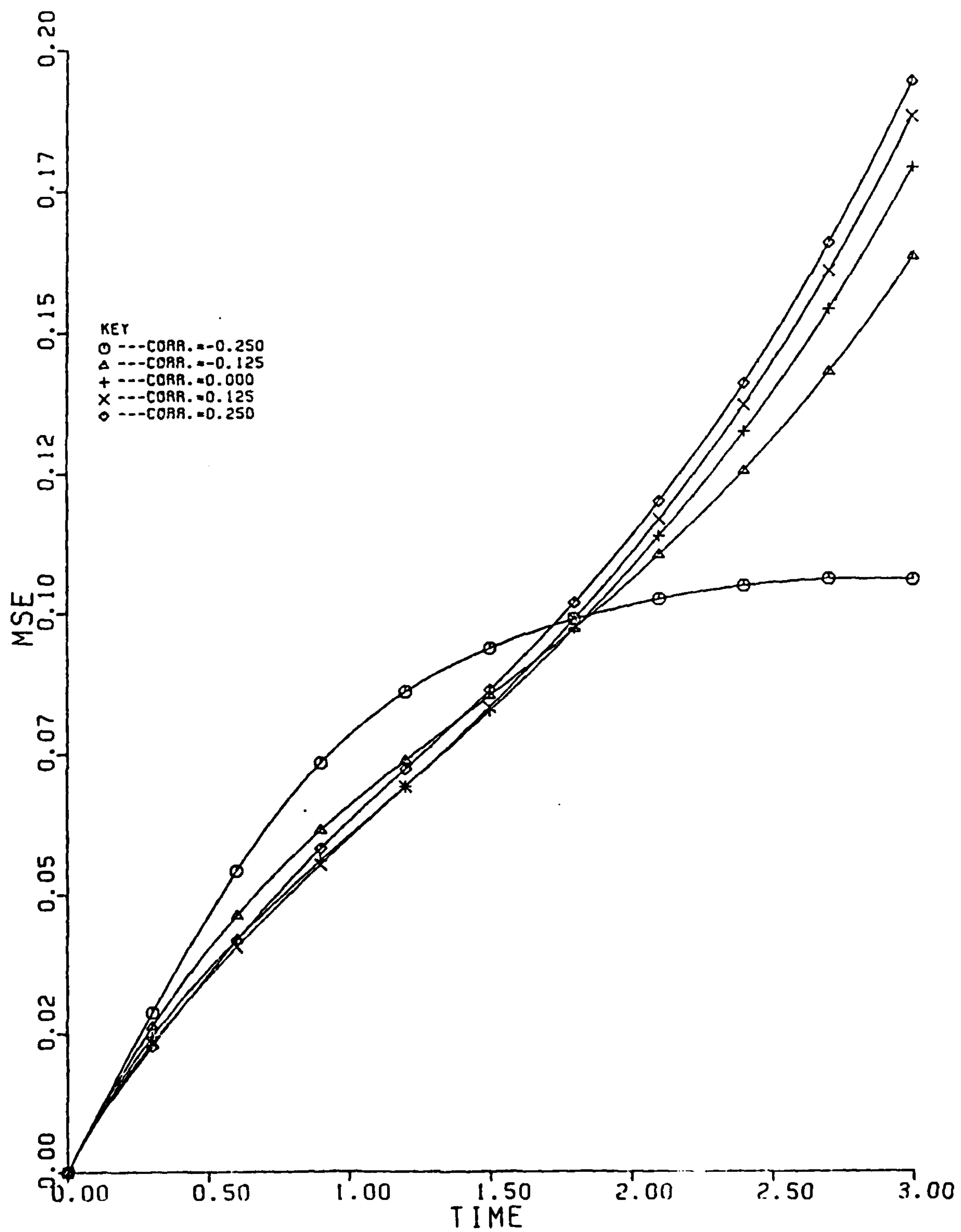
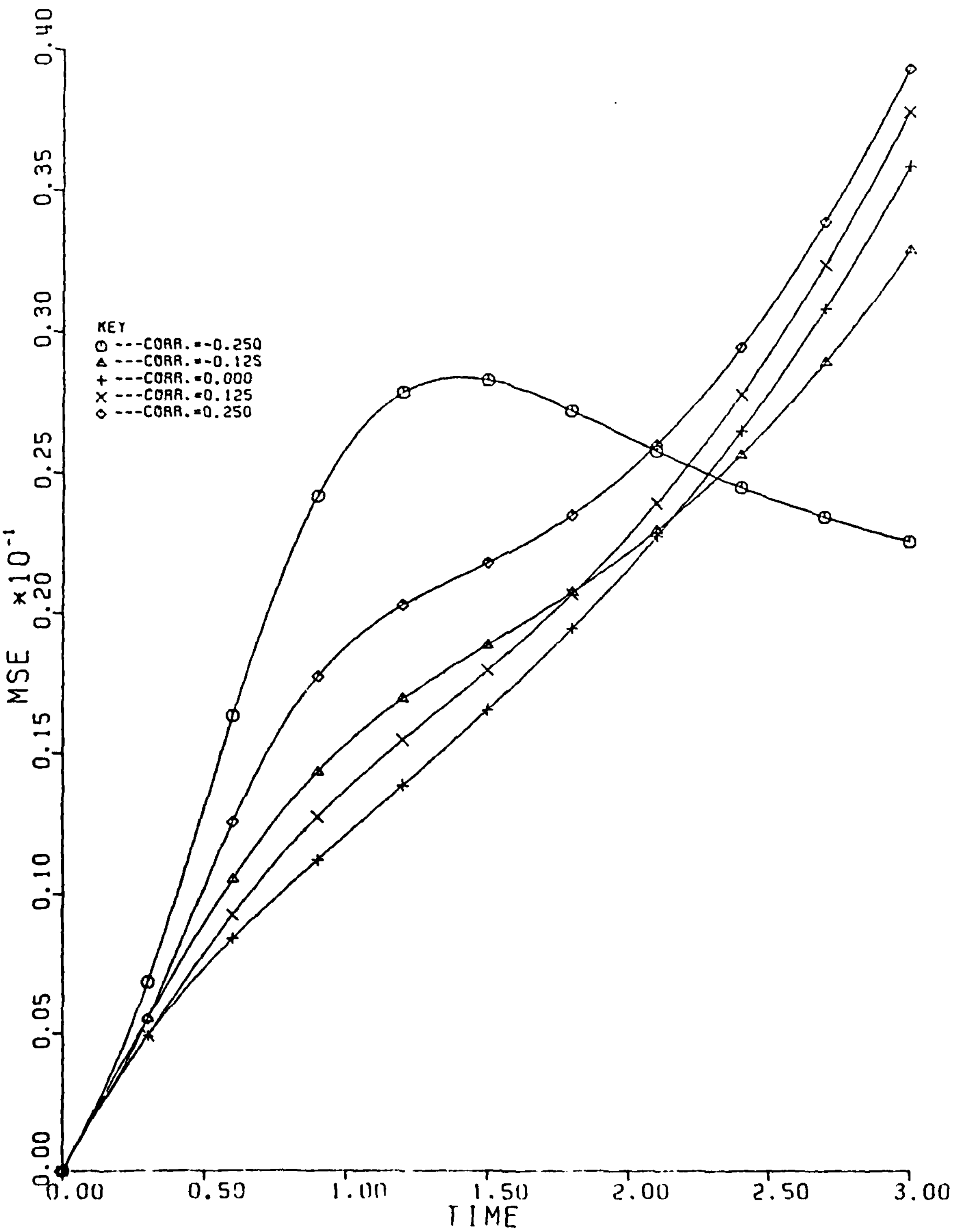


Figure 6B. MSE of Kaplan-Meier Estimate, $\tilde{\tilde{F}}_1(t)$, λ_1=1., λ_2=1.5, n=50



Note that in the case of equal component lifetimes, $\lambda_1=\lambda_2=\phi$ the bias determined from (4.12) and (4.14) simplifies to

$$B(\tilde{\bar{F}}_i(t)) = e^{-\phi t}\left\{\left[1+4\ (1-e^{-\phi t})^2\right]^{\frac{1}{2}}-1\right\}. \quad (4.16)$$

In the general case, the integral in (3.11) needs to be evaluated numerically. The bias of the Kaplan-Meier estimator was calculated for various values of λ_i and ρ. A representative plot of the bias appears in Figure 5, where $\lambda_1=1, \lambda_2=1.5$, and $|\rho|=0$, .125, .250. It is apparent that the bias is largest for values of t in the neighborhood of an interval which captures the mean component lifetimes. The absolute value of the bias ranges from 0 to .11, in this example.

$MSE(\tilde{\bar{F}}_i(t))$ was calculated for various values of λ_i, n, and ρ. Its magnitude is typified in Figures 6A and 6B, where $\lambda_1=1, \lambda_2=1.5$ and n=10, 50, respectively. For $\lambda_1=1, \lambda_2=1.5$, and $n=\infty$ $MSE(\tilde{\bar{F}}\ (t))$ may be obtained by squaring (4.14) or by squaring the ordinate values in Figure 5. The mean squared error of the Kaplan-Meier estimator may be quite large for small sample size n and moderately large for "large" ρ, the former being a more crucial factor than the later.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the referees for their constructive comments.

This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-82-0307 and the Graduate School, The Ohio State University.

REFERENCES

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing Holt Rinehart, and Winston.

Basu, A.P. (1981). Identifiability problems in the theory of competing and complementary risks - a survey. Statistical Distribution in Scientific Work (Taillie, Patil, and Baldesaari, Eds.), Dorerecht, Holland: Reidel Publishing Co., 335-348.

Birnbaum, Z.W. (1979). On the Mathematics of Competing Risks. U.S. Dept. of HEW. Publication No. 79-1351.

Boardman, T.J. and Kendell, P.J. (1970). Estimation in compound exponential failure models. Technometrics 12, 891-900.

Chao, Anne (1981). Approximate mean squared errors of estimators of reliability for k-out-of-m system in the independent exponential case. J. Amer. Statist. Assoc. 76: 720-724.

Easterling, R.G. (1980). Book review of Z.W. Birnhaum monograph. Technometrics 22, 131-132.

Gumbel, E.J. (1960). Bivariate exponential distributions. J. Amer. Statist. Assoc. 55, 698-707.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. New York, Wiley.

Klein, John P. and Moeschberger, M.L. (1983). Asymptotic bias of the product limit estimator under dependent competing risks. Submitted for publication.

Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.

Lagakos, S.W. and Williams J.S. (1978). Models for censored survival analysis: A cone clas of variable-sum models. Biometrika 65: 181-189.

Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. New York, Wiley.

Mann, N.R. and Grubbs, F.E. (1974). Approximately optimum confidence bounds for system reliability based on component test data. Technometrics 16, 335-347.

Mann, N.R. (1974). Simplified expressions for obtaining approximately optimum system-reliability confidence bounds from exponential subsystem data. J. Amer. Statist. Assoc. 69, 492-495.

Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974). Methods for the Statistical Analysis of Reliability and Life Data. Wiley, New York.

Mendenhall, W. and Lehman, E.H., Jr. (1960). An approximation to the negative moments of the positive binomial useful in life testing. Technometrics 2, 227-242.

Miyamura, T. (1982). Estimating component failure rates from combined component and systems data: exponentially distributed component lifetimes. Technometrics 24: 313-318.

Moeschberger, M.L. and David, H.A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-933.

Peterson, A.V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73: 11-13.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. 72: 20-22.

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: Constant-sum and variable-sum models. Biometrika 64: 215-224.

APPENDIX B

CONSEQUENCES OF ASSUMING INDEPENDENCE IN A BIVARIATE EXPONENTIAL SERIES SYSTEM

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

30 September, 1983

Annual Report for Period 1 September, 1982 - 30 September, 1983

AFOSR - NM
Building 410
Bolling Air Force Base
Washington DC 20332

AFOSR - PKZ
Building 410
Bolling Air Force Base
Washington DC 20332

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Consequences of Assuming Independence in a Bivariate Exponential Series System		5. TYPE OF REPORT & PERIOD COVERED Annual Sept. 1 1982 to Sept. 30 1983
		6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763265
7. AUTHOR(s) John P. Klein and M. L. Moeschberger		8. CONTRACT OR GRANT NUMBER(s) AFOSR-82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington DC 20332		12. REPORT DATE 30 Sept. 1983
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington DC 20332		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited; approved for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

System Reliability; Competing Risks; Marshall-Olkin Bivariate Exponential; Series System

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A common assumption made in modeling and analyzing data from series system is that the component lives are statistically independent. This study investigates the magnitude of the errors one may incur by erroneously assuming the component lifetimes have independent exponential distribution when in fact the lifetimes follow the bivariate exponential distribution of Marshall and Olkin (1967).

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AUTHOR'S FOOTNOTE

John P. Klein is an Assistant Professor, Department of Statistics and M. L. Moeschberger is an Associate Professor, Department of Preventive Medicine, both at The Ohio State University. Their research was supported under Contract AFOSR-82-0307 for the Air Force Office of Scientific Research.

ABSTRACT

A common assumption made in modeling and analyzing data from series systems is that the component lives are statistically independent. This study investigates the magnitude of the errors one may incur by erroneously assuming the component lifetimes have independent exponential distributions when in fact the lifetimes follow the bivariate exponential distribution of Marshall and Olkin (1967).

Key Words: System Reliability; Competing Risks; Marshall-Olkin Bivariate Exponential; Series System.

1. INTRODUCTION

Consider a two component system where the components are linked in series. Such a system will function properly provided both components function properly. Suppose that, if the two components are tested separately, the respective times to failure are, X_1, X_2. A common assumption made in life testing is that these component failure times follow exponential distributions (cf. Epstein and Sobel (1954), Boardman and Kendall (1970), Mann (1974), Chao (1981), etc.).

Often, when modeling a system linking two components in series, the assumption that X_1, X_2 are statistically independent is invoked. In general such an assumption is not testable due to the identifiability dilemma (see Basu and Klein (1983) for a discussion and references). This assumption is important in both theoretical modeling of the resulting system and in obtaining estimators of component parameters from data collected on series systems.

The purpose of this study is to investigate the effects of departures from this independence assumption on both the modeling and estimation problem. The specific form of departure from independence which we shall assume is that the joint distribution of X_1, X_2 is the bivariate exponential distribution of Marshall and Olkin (1967). This distribution has joint survival function

$$P(X_1>x_1,\ X_2>x_2) = \exp(-\lambda_1 x_1-\lambda_2 x_2-\lambda_{12}\max(x_1,x_2))$$

$$\text{for } \lambda_1,\ \lambda_2 > 0,\ \lambda_{12} \geq 0,\ x_1,\ x_2 > 0. \qquad (1.1)$$

The marginal distributions of X_1 and X_2 are exponential with parameters $\Theta_i = \lambda_i + \lambda_{12}$, $i = 1, 2$. This distribution can be derived from both a fatal and non-fatal shock model. It has a singular component when λ_{12} is nonzero which reflects the possibility of the system receiving a shock of sufficient magnitude to simultaneously destroy both components. The distribution reduces to that of independent exponentials when $\lambda_{12} = 0$.

In Section 2 we examine the effects of assuming the components to have independent life-lengths in modeling series systems, on the system reliability, system mean life and crude system life, when in fact the distribution of lifetimes follows the bivariate exponential distribution (1.1). In Section 3 we investigate the effects of the independence assumption on estimating component parameters when in fact the bivariate exponential is the correct model.

2. MODELING SERIES SYSTEMS

2.1 System Reliability

Suppose that, based on testing each component separately, an investigator knows that the marginal distributions of the component lifetimes are given by $P(X_i > x) = \exp(-(\lambda_i+\lambda_{12})x)$, $i = 1, 2$. Also, suppose the investigator believes, for the purpose of modeling system life, that the components are statistically independent when in fact the joint distribution is the bivariate exponential (1.1). Under the assumption of independence the system reliability at a mission time t is give by

$$P(\min(X_1, X_2) > t \mid \text{Independence}) = \bar{F}_I(t) = \exp(-(\lambda+\lambda_{12})t),$$

where $\lambda = \lambda_1+\lambda_2+\lambda_{12}$, while under the bivariate exponential model we have

$$P(\min(X_1, X_2) > t \mid \text{BVE}) = \bar{F}_D(t) = \exp(-\lambda t).$$

The amount of error made in assuming independence is

$$\bar{F}_D(t) - \bar{F}_I(t) = \exp(-\lambda t)(1-\exp(-\lambda_{12}t)) \qquad (2.1.1)$$

which is always non-negative so that the system is more reliable then one would expect under independence.

For λ_1, λ_2, λ_{12} fixed the maximum difference between $\bar{F}_D(t)$ and $\bar{F}_I(t)$ occurs at $t = \lambda_{12}^{-1} \ln[1+\lambda_{12}/\lambda]$ and the magnitude of this difference is $\rho(1+\rho)^{-1/\rho}/(1+\rho)$ where $\rho = \lambda_{12}/\lambda$ is the correlation between X_1 and X_2. This maximum difference approaches 1/4 as ρ increases to one.

Alternatively, an investigator modeling system life, based on component knowledge, may wish to assess the error (2.1.1) at some fixed mission time. Consequently, for λ_1, λ_2, and $t > 0$ fixed, the maximum deviation between $\bar{F}_D(t)$ and $\bar{F}_I(t)$ occurs at $\lambda_{12} = (\ln 2)/t$ and this maximum difference is $.25 \exp(-(\lambda_1+\lambda_2)t)$. In the range of values of t where $\bar{F}_I(t)$ is less than 0.5 the magnitude of this difference is $\bar{F}_I(t)$ itself so that the true system reliability is twice as large as that predicted using the erroneous assumption of independence.

2.2 System Mean Life

Under the assumption of independent component lifetimes the mean time to failure of the series system is $\mu_I = (\lambda+\lambda_{12})^{-1}$ while under the correct bivariate exponential model the mean system life is $\mu_D = \lambda^{-1}$. Clearly $\mu_D - \mu_I = \lambda_{12}/\lambda(\lambda+\lambda_{12}) = \rho\mu_I$

which is always non-negative. The largest difference between μ_D and μ_I occurs when $\lambda_{12} = (\lambda_1+\lambda_2)/\sqrt{2}$ and the magnitude of this difference is $(3-2\sqrt{2})/(\lambda_1+\lambda_2) \doteq .1716/(\lambda_1+\lambda_2)$. Again as in the case of system reliability one may be appreciably misled when ρ is not zero.

2.3 $P(X_1 < X_2)$

Investigators working with series systems would like to know the value of $P(X_1 < X_2)$ so that they can determine the dominant cause of system failure and hence concentrate their efforts on improving the reliability of that component. Under the assumption of independence we have $\Pi_I = P(X_1 < X_2 | \text{Independence}) = (\lambda_1+\lambda_{12})/(\lambda+\lambda_{12})$ while under the dependence model we have $\pi_D = P(X_1 < X_2 | \text{BVE}) = \lambda_1/\lambda$. The difference between π_I and π_D is $\lambda_{12}(\lambda-\lambda_1)/(\lambda(\lambda+\lambda_{12})) = \rho(1-\pi_D)/(1+\rho)$ which is an increasing non-negative function of ρ with a supremum of 1/2 as ρ goes to unity. Similar results hold for $P(X_1 > X_2)$.

2.4 Crude System Life

An observable quantity in series systems is the crude system life given by $Q_j(t) = P(\min(X_1, X_2) > t, \min(X_1,X_2)=X_j)$, $j = 1, 2$. For the jth component this is the observable survival function for systems where the jth component fails first. Under the assumption of independence we have

$Q_j(t|\text{Independence}) = Q_j^{(I)}(t) = ((\lambda_j+\lambda_{12})/(\lambda+\lambda_{12}))\exp(-(\lambda+\lambda_{12})t)$

while under the correct bivariate exponential model we have

$Q_j(t|\text{BVE}) = Q_j^{(D)}(t) = (\lambda_1/\lambda)\exp(-\lambda t)$. Hence we have

$$Q_j^{(I)}(t)-Q_j^{(D)}(t) = \exp(-\lambda t)\left\{\frac{\lambda_j}{\lambda} - \frac{(\lambda_j+\lambda_{12})}{(\lambda+\lambda_{12})}\exp(-\lambda_{12}t)\right\}.$$

For λ_{12} not zero, the crude survival curve under independence crosses the curve computed under dependence once from above at $t_0 = -\lambda_{12}^{-1}\ln[\lambda_j(\lambda+\lambda_{12})/(\lambda(\lambda_j+\lambda_{12}))]$. In the range $t < t_0$ the maximum difference between $Q_j^{(I)}(t)$ and $Q_j^{(D)}(t)$ occurs at the origin and is the same as the difference between the respective probabilities that the jth component fails first. In the range $t > t_0$ the maximum deviation between the two curves is at $t_1 = -\lambda_{12}^{-1}\ln[\lambda_j/(\lambda_j+\lambda_{12})]$ and this maximum deviation is

$$\left[\frac{\lambda_j}{\lambda_j+\lambda_{12}}\right]^{\lambda/\lambda_{12}}\left[\frac{\lambda_{12}\ \lambda_j}{\lambda(\lambda+\lambda_{12})}\right] = \frac{\rho\ \pi_{Dj}}{(1+\rho)}\cdot\left(\frac{\pi_{Dj}}{\pi_{Dj}+\rho}\right)^{1/\rho}$$

where π_{Dj} is the probability the jth component fails first under the bivariate exponential model.

3. ESTIMATING COMPONENT RELIABILITY

A second area where the assumption of independence plays a key role is in estimating component parameters from data collected on series systems. Continuing in the same vein as in Section 2, we assume the components have exponentially distributed lifetimes with unknown failure rates Θ_1, Θ_2. We put n systems on test and for each system we observe the failure time, t_i, and an indicator variable which tells us which component caused the system failures.

If the data is from systems whose components are independent then the sufficient *statistics for* Θ_i are $T = \Sigma t_i$ and M_1 the number of systems where the first component failed first. The maximum likelihood estimator of Θ_j is $\hat{\Theta}_1 = M_1/T$ and $\hat{\Theta}_2 = (n-M_1)/T$. David and Moeschberger (1978) show that the estimators $\tilde{\Theta}_j = ((n-1)/n)\hat{\Theta}_j$ are the best unbiased estimators of Θ_j and that for $n > 2$ the variance of $\tilde{\Theta}_j$ is $V(\tilde{\Theta}_j) = [(n-1)\Theta_j(\Theta_1+\Theta_2)+\Theta_j^2]/(n(n-2))$.

Now suppose an investigator carrys out the above analysis, believing the component lifetimes are independent, when in fact the joint distribution of the component lifetimes is bivariate exponential. If the data is from a bivariate exponential series system there is a positive probability that the two components failed simultaneously. In such a case, suppose the investigator, due to his strong belief in the independence assumption where such simultaneous failures are impossible, records this occurrence as a failure from component 1 with probability ϕ and a failure from component 2 with a probability $1-\phi$. Such action is done independently of the system under consideration and this probability is constant from one system to another. In this setting the parameters of interest are $\Theta_j = \lambda_j+\lambda_{12}$, $j = 1, 2$. For the ith system, let X_{1i}, X_{2i} be the unobservable component lifetimes. The recorded statistics are now $T = \Sigma \min(X_{1i}, X_{2i})$ and $M_1 = \Sigma\psi(X_{1i}, X_{2i})$ where

$$\psi(X_{1i}, X_{2i}) = \begin{cases} 1 \text{ if } X_{1i} < X_{2i} \text{ or } X_{1i} = X_{2i} \text{ and } Y_i = 1 \\ 0 \qquad\qquad \text{otherwise} \end{cases}$$

where the Y_i's are independent and identically distributed Bernoulli random variable with $P(Y_i = 1) = \phi$.

Under the bivariate exponential distribution T has a gamma distribution with parameters n and λ so $E(T^{-1}) = \lambda/(n-1)$, $E(T^{-2}) = \lambda^2/[(n-1)(n-2)]$ if $n > 2$. Also $E(M_1) = n(\lambda_1+\phi\lambda_{12})/\lambda$ and $E(M_1^2) = n(\lambda_1+\phi\lambda_{12})/\lambda+(n-1)\,n[\lambda_1+\phi\lambda_{12}]^2/\lambda^2$. Also T is independent of M_1. Thus the bias of $\tilde{\Theta}_j$, under this model is $B(\tilde{\Theta}_1) = E(\tilde{\Theta}_1-\Theta_1) = \lambda_{12}(1-\phi)$ and $B(\tilde{\Theta}_2) = \lambda_{12}\phi$. In both cases the bias is an increasing function of the correlation ρ so that the more correlated the components the greater the magnitude of the estimation error. The respective mean squared errors are

$$MSE(\tilde{\theta}_1) = [(n-1)\lambda(\lambda_1+\phi\lambda_{12})+(\lambda_1+\phi\lambda_{12})^2]/[n(n-2)]+(1-\phi)^2\lambda_{12}^2$$

and

$$MSE(\tilde{\theta}_2) = [(n-1)\lambda(\lambda_2+(1-\phi)\lambda_{12})+(\lambda_2+(1-\phi)\lambda_{12})^2]/[n(n-2)]+\phi^2\lambda_{12}^2$$

which are quadratic functions of ρ. As n goes to infinity both mean squared errors tend to nonzero constants so the estimators are inconsistent.

REFERENCES

BASU, ASIT P. and KLEIN, JOHN P. (1983). "Some Recent Results in Competing Risks," Proceedings of IMS Survival Conference (to appear).

BOARDMAN, T. J. and KENDELL, P. J. (1970). "Estimation in Compound Exponential Failure Models," Technometrics 12, 891-900.

CHAO, ANNE (1981). "Approximate Mean Squared Errors of Estimators of Reliability for k-out-of-m Systems in the Independent Exponential Case," Journal of the American Statistical Association 76, 720-724.

DAVID, H. A. and MOESCHBERGER, M. L. (1978). The Theory of Competing Risks, New York: MacMillan Publishing Co., Inc.

EPSTEIN, B. and SOBEL, M. (1954). "Some Theorems Relevant to Life Testing From an Exponential Distribution," Annals of Mathematical Statistics, 25, 383-381.

MANN, N. R. (1974). "Simplified Expressions for Obtaining Approximately Optimum System Reliability Confidence Bounds from Exponential Subsystem Data," Journal of the American Statistical Association 69, 492-495.

MARSHALL, A. W. and OLKIN, I. (1967). "A Multivariate Exponential Distribution," Journal of the American Statistical Association, 62, 30-40.

APPENDIX C

ASYMPTOTIC BIAS OF THE PRODUCT LIMIT ESTIMATOR UNDER DEPENDENT COMPETING RISKS

John P. Klein
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

M. L. Moeschberger
Department of Preventive Medicine
The Ohio State University
Columbus, Ohio 43210

30 September, 1983

Annual Report for Period 1 September, 1983 - 30 September, 1983

AFOSR - NM
Building 410
Bolling Air Force Base
Washington DC 20332

AFOSR - PKZ
Building 410
Bolling Air Force Base
Washington DC 20332

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Asymptotic Bias of the Product Limit Estimator under Dependent Competing Risks	5. TYPE OF REPORT & PERIOD COVERED Annual Sept. 1 1982 to Sept. 30, 1983
	6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763265
7. AUTHOR(s) John P. Klein and M. L. Moeschberger	8. CONTRACT OR GRANT NUMBER(s) AFSOR-82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR-NM Building 410 Bolling AFB, Washington DC 203323	12. REPORT DATE 30 Sept. 1983
	13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington DC 20332	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited; approved for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Competing risks, product limit estimator, dependent risks, consistency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A common assumption mades in analyzing competing risk experiments is that the risks are stochastically independent. Under that assumption the product limit estimator is a consistent estimator of the marginal survival function. We show that when the risks are not independent the product limit estimator converges, with probability one, to a survival function which may not be the same as the marginal survival function of interest.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

A common assumption made in analyzing competing risk experiments is that the risks are stochastically independent. Under that assumption the product limit estimator is a consistent estimator of the marginal survival function. We show that when the risks are not independent the product limit estimator converges, with probability one, to a survival function which may not be the same as the marginal survival function of interest.

Key Words and Phrases: Competing risks, product limit estimator, dependent risks, consistency.

*John P. Klein is Assistant Professor, Department of Statistics, The Ohio State University, Columbus, Ohio 43210. M. L. Moeschberger is Associate Professor, Department of Preventive Medicine, The Ohio State University, Columbus, Ohio 43210.
Both Professors Klein and Moeschberger acknowledge support for this research under Contract AFOSR-82-0307 for the Air Force Office of Scientific Research.

1. INTRODUCTION

Competing risks arise in a wide range of life testing problems. Typical areas of application are the study of series systems in the engineering sciences and biological systems in the medical sciences. An important area of application is the analysis of censored data where some systems or individuals are lost or withdrawn from a study prior to observing the endpoint of interest. Competing risks are often modeled by a vector $\underset{\sim}{T} = (T_1, \ldots, T_p)$ of nonnegative random variables representing the potential times to failure from each of the p causes. We cannot observe $\underset{\sim}{T}$ directly but instead we see the system failure time $Y = \min(T_i, i = 1, \ldots, p)$ and the failure pattern $\xi(\underset{\sim}{T}) = I$ such that $Y = T_i$ for $i \in I$ and $Y < T_i$ $i \notin I$, where $I \varepsilon \mathcal{I}$ the set of all subsets of $\{1, \ldots, p\}$. Based on this information we wish to estimate the marginal survival probabilities $S_I(t) = P(\min(T_i, i \varepsilon I) > t)$, $t > 0$, $I \varepsilon \mathcal{I}$.

A common assumption made in analyzing competing risk experiments is that the T_i's are independent random variables. Such an assumption is not testable due to the identifiability dilemma (see Basu (1981)). Under an assumption of independent risks a consistent estimator of $S_I(t)$ is the Kaplan and Meier (1958) product limit estimator. In Section 2 we show that, if the risks are dependent, then the product limit estimator may be an inconsistent estimator of $S_I(t)$. The quantity to which this estimator converges is obtained so that one may investigate the estimator's robustness to departures from independence. In Section 3 we illustrate such robustness considerations for some well-known bivariate exponential distributions.

2. INCONSISTENCY OF THE PRODUCT LIMIT ESTIMATOR

The Kaplan and Meier product limit estimator is constructed as follows. Let $0 = Y_{(0)} \leq Y_{(1)} \leq \cdots \leq Y_{(n)}$ denote the ordered system failure times of n systems put on test. The product-limit estimator of $S_I(t)$ is

$$\hat{S}_I(t) = \Pi_i[(n-i)/(n-i+1)] \tag{2.1}$$

where the product is over the ranks i of those ordered observations $Y_{(i)}$, $1 \leq i \leq n$, such that $Y_{(i)} \leq t$ and $Y_{(i)}$ corresponds to a death from the simultaneous cause(s) $j \in J$, $J \cap I \neq \emptyset$. $\hat{S}_I(t)$ is undefined for $t > Y_{(n)}$ if the largest failure time corresponds to causes in J where $J \cap I = \emptyset$.

If the assumption of independence is correct and the crude probability functions defined by $F(t, I) = P(Y > t, \xi(\underset{\sim}{T}) = I)$ have no common discontinuities then Langberg, Proschan and Quinzi (1981) [LPQ(1981)] have shown that the product limit estimator is consistent. They also show that if the F(t,I)'s have no common discontinuities then for a very particular form of dependence structure the product limit estimator is consistent. We note in the following theorem, that their results can be used to study the robustness of the product limit estimator to departures from independence and that, in general, if the risks are dependent then the product limit estimator (2.1) is inconsistent.

THEOREM 1. Let $\underset{\sim}{T} = (T_1, \ldots, T_p)$ be a vector of non-negative random variables with system life $Y = \min(T_1, \ldots, T_p)$ and failure pattern

$$\xi(\underset{\sim}{T}) = I, \text{ if } Y = T_i,\ i \varepsilon I \text{ and } Y < T_i,\ i \not\varepsilon I \qquad (2.2)$$
$$\emptyset, \quad \text{otherwise}$$

Define $\bar{F}(t) = P(Y > t)$, $F(t, I) = P(Y < t, \xi(\underset{\sim}{T}) = I)$, $I \varepsilon \mathcal{I}$, $t > 0$, and let $\alpha(\bar{F}) = \{t: \bar{F}(t) > 0\}$ be the support of $\bar{F}$. For $I \varepsilon \mathcal{I}$ define $\mathcal{I}_I = \{J \varepsilon \mathcal{I}: J \quad I \neq \emptyset\}$. Based on a random sample of size n let $\hat{S}_{I,n}(t)$ be the product limit estimator (2.1). If the functions $F(\cdot, I)$ have no common discontinuities on $[0, \alpha(\bar{F})]$ then

$$\hat{S}_{I,n}(t) \to \prod_{J \varepsilon \mathcal{I}_I} \bar{G}_J(t) \quad \text{a.s.} \qquad (2.3)$$

where

$$\bar{G}_J(t) = \prod_{a \leq t} [\bar{F}(a)/\bar{F}(a^-)] \exp[-\int_0^t (dF^c(.,J)/\bar{F})],\ 0 \leq t \leq \alpha(\bar{F}). \qquad (2.4)$$

where the product is over the set of discontinuities of $F(\cdot, J)$ and $F^c(\cdot, J)$ is the continuous part of $F(\cdot, J)$.

PROOF. The proof follows directly by applying the results of Langberg, Proschan and Quinzi (1978) [LPQ(1978)] and LPQ(1981). By Theorem 4.1 of LPQ(1978) $T =_{LP} H$ where H is a vector of $(2^p - 1)$ independent components indexed lexicographically by $I \varepsilon \mathcal{I}$ with $P(H_I > t) = \bar{G}_I(t)$ given by (2.4).

Let $\underset{\sim}{T}_i = (T_{1i}, \ldots, T_{pi})$, $i = 1, \ldots, n$ be independent and identically distributed as $\underset{\sim}{T}$. Replace $\bar{F}$ and $F(\cdot,J)$ in (2.4) by their empirical counterparts $\hat{\bar{F}}_n(t) = \sum_{i=1}^{n} \chi\{Y_i > t\}/n$ and $\hat{F}_n(t,J) = \sum_{i=1}^{n} \chi\{Y_i \leq t, \xi(\underset{\sim}{T}_i) = J\}/n$, to obtain $\hat{\bar{G}}_{I,n}(t)$. Here $\chi(A)$ is the indicator function of the set A.

By Theorem 4.7 of LPQ[1981] $\hat{\bar{G}}_{I,n}(t) \to \bar{G}_J(t)$ a.s. for $J \in \mathcal{I}$. Routine algebraic minipulation shows that $\hat{S}_{I,n}(t) = \prod_{J \in \mathcal{I}_I} \hat{\bar{G}}_{J,n}(t)$ so the result now follows $\square$.

In general, as seen in the examples in the following section, $\prod_{J \in \mathcal{I}_I} \bar{G}_J(t) \neq S_I(t)$ so that an investigator may be seriously misled by incorrectly assuming that the component lifetimes are independent. This has been noticed by Fisher and Kanarek (1974) in the problem of analyzing clinical trials with censored data. Theorem 1 allows an investigator to quantify the effects of the independence assumption by computing the right hand side of (2.3) for some plausible dependent models.

LPQ(1981) have shown that for a special type of dependence the estimator $\hat{S}_{I,n}(t)$ is consistent. We state their result, without proof, as a corollary.

COROLLARY 1. Assume that the conditions of Theorem 1 hold. Then $\hat{S}_{I,n}(t) \to S_I(t)$ a.s. if and only if the following two conditions hold.

i) $S_I(a)/S_I(a^-) = \bar{F}(a)/\bar{F}(a^-)$, a discontinuity point of $\sum F(G,J)$, where the sum is over $J \varepsilon \mathcal{I}_I$

$= 1$, otherwise.

ii) $P(\min(T_i,\ i\varepsilon I') \geq t \mid \min(T_i,\ i\varepsilon I)=t)$

$= P(\min(T_i,\ i\varepsilon I')>t \mid \min(T_i,\ i\varepsilon I)>t)$

where I' is the compliment of $I \varepsilon \mathcal{I}$.

EXAMPLES

In this section we present some representation examples of the use of Theorem 1 in determining the effects, on estimating marginal survival, of the independence assumption for some bivariate exponential life distributions. Let (T_1,T_2) denote the time to failure from components 1 and 2, respectively, in a series system. Let $\bar{F}(t_1, t_2)$ be the joint survival function of (T_1, T_2) and $S_i(t) = P(T_i > t)$, i = 1, 2 the marginal survival functions. Let $\hat{S}_i(t)$ be the estimator (2.1) of $S_i(t)$, and let (t), J={1},{2},{1,2} be given by (2.4). Then $\hat{S}_i(t) \to \bar{G}_{\{i\}}(t)\bar{G}_{\{1,2\}}(t)$ a.s by Theorem 1, if the functions F(t, I) have no common discontinuities. Note that if $P(T_1 = T_2) = 0$ then $\bar{G}_{\{1,2\}}(t)=1$ for all t. We now give some examples

EXAMPLE 1. (Block and Basu (1974))

Let $\bar{F}(t_1, t_2) = [\lambda/(\lambda_1+\lambda_2)]\exp(-\lambda_1 t_1-\lambda_2 t_2-\lambda_{12}\max(t_1,t_2))]$
$- [\lambda_n/(\lambda_1+\lambda_2)]\exp(-\lambda\max(t_1,t_2))$, for
$t_1, t_2 > 0$,

$\lambda_1, \lambda_2 > 0 \quad \lambda_{12} \geq 0, \lambda = \lambda_1 + \lambda_2 + \lambda_{12}$.

Here $S_i(t) = \frac{\lambda}{(\lambda_1+\lambda_2)} \exp(-(\lambda_i+\lambda_{12})t) - \frac{\lambda_{12}}{(\lambda_1+\lambda_2)} \exp(-\lambda t)$,

$t > 0$, but

by theorem 1

$\hat{S}_i(t) \to \exp(-\frac{\lambda_i\lambda}{(\lambda_1+\lambda_2)} t)$, a.s. $t > 0$, $i = 1, 2$.

EXAMPLE 2. Gumbel (1960)

Let $\bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} t_1 t_2)$ $t_1, t_2 > 0$,
$\lambda_1+\lambda_2 > 0$, $\lambda_{12} \geq 0$.

Here $S_i(t) = \exp(-\lambda_i t)$, $t \geq 0$ but by Theorem 1

$\hat{S}_i(t) \to \exp(-\lambda_i t - \lambda_{12} t^2/2)$, a.s. for $t \geq 0$, $i = 1, 2$.

EXAMPLE 3. Gumbel (1960)

Let $\bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1-\lambda_2 t_2)[1+\lambda_{12}$
$-\lambda_{12}(\exp(-\lambda_1 t_1) + \exp(-\lambda_2 t_2))]$,
$+ \lambda_{12}\exp(-\lambda_1 t_1-\lambda_2 t_2)]$
$t_1, t_2 > 0$, $\lambda_1, \lambda_2 > 0$, $\lambda_{12} \geq 0$.

Here $S_i(t) = \exp(-\lambda_i t)$ but by Theorem 1

$$\hat{S}_i(t) \to \exp\{-\lambda_i \int_0^t \frac{[1+\lambda_{12}(1-e^{-\lambda_j t})(1-2e^{-\lambda_i t})]}{[1+\lambda_{12}(1-e^{-\lambda_i t})(1-e^{-\lambda_j t})} dt \quad \text{where}$$

j is the compliment of i in {1, 2}.

<u>EXAMPLE 4</u>. Marshal-Olkin (1967)

Let $\bar{F}(t_1, t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2))$,

$t_1, t_2 > 0,\ \lambda_1, \lambda_2 \geq 0,\ \lambda_{12} \geq 0.$

Here $S_i(t) = \exp(-(\lambda_i + \lambda_{12})t)$. In this case the conditions of Corollary 1 are met so $\hat{S}_i(t) \to S_i(t)$ a.s.

REFERENCES

Basu, A. P. (1981), "Identifiability problems in the theory of competing and complimentary risks - a survey," in Statistical Distributions in Scientific Work, eds. Taillies, Patil and Baldesaari, Dorerecht, Holland: Reidel Publishing Company, 335-348.

Block, H. W. and Basu, A. P. (1974), "A continuous bivariate exponential extension," Journal of the American Statistical Association, 69, 1031-1037.

Fisher, L. and Kanarek, P. (1974), "Presenting censored survival data when censoring and survival times may not be independent," in Reliability and Biometry: Statistical Analysis of Lifelength, Eds. Proschan and Serfling, Philadelphia, Society for Industrial and Applied Mathematics, 303-326.

Gumbel, E. J. (1960), "Bivariate exponential distributions," Journal of the American Statistical Association, 55, 698-707.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations," Journal of the American Statistical Association, 53, 457-481.

Langberg, N., Proschan, F. and Quinzi, A. J. (1978), "Converting dependent models into independent ones, preserving essential features," Annals of Probability, 6, 174-181.

__________________ (1981), "Estimating dependent life lengths with applications to the theory of competing risks," Annals of Statistics, 9, 157-167.

Marshal, A. W. and Olkin, I. (1967), "A multivariate exponential distribution," Journal of the American Statistical Association, 62, 30-44.

APPENDIX D

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Test for Independence Based on Data from a Series System		5. TYPE OF REPORT & PERIOD COVERED Annual Sept 1 1982 to Sept. 30 1983
		6. PERFORMING ORG. REPORT NUMBER 714837 MPN 763265
7. AUTHOR(s) John P. Klein		8. CONTRACT OR GRANT NUMBER(s) AFOSR-82-0307
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Ohio State University Research Foundation 1314 Kinnear Road Columbus, Ohio 43212		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFODR-NM Building 410 Bolling AFB, Washington DC 20332		12. REPORT DATE 30 Sept. 1983
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) AFOSR-PKZ Building 410 Bolling AFB Washington DC 20332		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Series Systems; Test for Independence; Kendall's Tau; Exponential Distribution.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of testing for independence of the component lifetimes when the components are linked in series is considered. To avoid the problem of nonidentifiability the marginal component lifetimes are assured to be known. In this setting a modified version of Kendall's Tau is proposed. This test statistic is obtained by replacing those component lifetimes which cannot be observed, due to system failure, by conditional probabilities computed under independence. A small scale simulation study of the power of this test shows the test has reasonable power for relatively small sample sizes.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

A TEST FOR INDEPENDENCE BASED ON DATA
FROM A SERIES SYSTEM

by

JOHN P. KLEIN
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

Abstract

The problem of testing for independence of the component lifetimes when the components are linked in series is considered. To avoid the problem of nonidentifiability the marginal component lifetimes are assumed to be known. In this setting a modified version of Kendall's Tau is proposed. This test statistic is obtained by replacing those component lifetimes which cannot be observed, due to system failure, by conditional probabilities computed under independence. A small scale simulation study of the power of this test shows the test has reasonable power for relatively small sample sizes.

Key Words: Series Systems; Test for Independence; Kendall's Tau; Exponential Distribution.

1. INTRODUCTION

A common assumption made in modeling series systems is that the component lifetimes are statistically independent. This independence assumption is also routinely made in analyzing data collected from series systems. Recently, Klein and Moeschberger (1983) and Moeschberger and Klein (1982) have shown that one may be appreciably misled by erroneously assuming the independence of component lifelengths in modeling and estimating the parameters of certain bivariate exponential series systems. Thus it is desirable to have a test of this assumption of independence based on data from series systems.

If no assumption about the underlying distribution of the component lifetimes is made it is impossible to test for independence due to the identifiability problem (see, e.g. Tsiatis (1978), Miller (1977) or Basu and Klein (1982)). That is, given any set of observable information, collected from a series system with dependent component lifetimes, there exists a series system with independent component lifetimes which yields the same set of observable information. Hence, any test of independence must incorporate some additional information about the underlying model to be able to detect departures from independence.

In Section 2 a modification of Kendall's (1938) test for independence is considered. This modification assumes that the marginal component life distributions are completely specified. This information could be obtained by testing each component separately, as if often done in the development stages of system design (see, e.g. Easterling and Prairie (1971), Mastran (1976), and Miyamura (1982)). In Section 3 a small scale simulation study shows that the test performs fairly well for relatively small sample sizes.

2. THE TEST PROCEDURE

Suppose that n two component series systems are put on test. Let X_i, Y_i denote the potential (unobservable) failure times of the first and second components of the i^{th} system. We are not allowed to observe (X_i, Y_i) directly, but instead we observe $T_i = \min(X_i, Y_i)$, the system failure time and $I_i = \begin{cases} 1 & \text{if } T_i = X_i \\ 0 & \text{if } T_i = Y_i \end{cases}$, the cause of the system failure.

Also suppose that the marginal survival functions of X_i and Y_i, $\overline{F}(x) = P(X_i > x)$ and $\overline{G}(y) = P(Y_i > y)$, $i = 1, \ldots, n$ are known.

If we could observe both X_i and Y_i then a test of independence, due to Kendall (1938), is to count the number of concordant pairs and the number of discordant pairs. A pair (X_i, Y_i), (X_j, Y_j) is concordant if $X_i - X_j$ and $Y_i - Y_j$ have the same sign and is discordant if these differences have different signs. The test statistic is then the number of concordant pairs minus the number of discordant pairs.

If the data comes from a series system then only T_i, I_i is observed. Suppose we consider a pair (T_i, I_i), (T_j, I_j) with $T_i < T_j$. If $I_i = 1$ and $I_j = 1$ then we know that $X_i = T_i < X_j = T_j$, and $X_i < Y_i$, $X_j < Y_j$. This pair would be concordant, regardless of the value of Y_j, if $T_i < Y_i < T_j$. If $Y_i > T_j$ concordance or discordance depends on the value of Y_j. Under the null hypothesis of independence, the conditional probability that the pair is concordant is $[\overline{G}(T_i) - \overline{G}(T_j)]/\overline{G}(T_i)$. When $I_i = 1$ and $I_j = 0$ then $T_i = X_i < Y_j = T_j$, $X_i < Y_i$, $Y_j < X_j$. Here if $T_i < Y_i < T_j$ the pair would be concordant and if $Y_i > T_j$ the pair would be discordant, whatever the value of X_j. Under independence the conditional probabilities of these two events are $[\overline{G}(T_i) - \overline{G}(T_j)]/\overline{G}(T_i)$ and $\overline{G}(T_j)/\overline{G}(T_i)$, respectively. Should

$I_i = 0$ similar probabilities, involving $\overline{F}$, could be obtained. This motivation suggests the following score function for $T_i < T_j$

$$\phi(T_i,I_i,T_j,I_j) = \begin{cases} [\overline{G}(T_i)-\overline{G}(T_j)]/\overline{G}(T_i) & \text{if } I_i = I_j = 1 \\ [\overline{F}(T_i)-\overline{F}(T_j)]/\overline{F}(T_i) & \text{if } I_i = I_j = 0 \\ [\overline{G}(T_i)-2\overline{G}(T_j)]/\overline{G}(T_i) & \text{if } I_i = 1, I_j = 0 \\ [\overline{F}(T_i)-2\overline{F}(T_j)]/\overline{F}(T_i) & \text{if } I_i = 0, I_j = 1 \end{cases} \tag{2.1}$$

and similarly for $T_i > T_j$.

The modified version of Kendall's test statistic is

$$\hat{\tau} = \sum_{1\leq i<j\leq n}\sum \phi(T_i,I_i,T_j,I_j)/\binom{n}{2}. \tag{2.2}$$

In the appendix we show, that under the null hypothesis of independence of the component lifetimes:

A) $E(\hat{\tau}) = 0$

B)
$$\begin{aligned} n(n-1)V(\hat{\tau}) &= \frac{2}{3}\int_{-\infty}^{\infty}\overline{G}(x)^2dF(x) + \frac{2}{3}\int_{-\infty}^{\infty}\overline{F}(x)^2dG(x) \\ &- 2\int_{-\infty}^{\infty}\overline{G}(x)^{-1}\int_{x}^{\infty}F(y)\overline{G}(y)^2dG(y)dF(x) \\ &- 2\int_{-\infty}^{\infty}\overline{F}(x)^{-1}\int_{x}^{\infty}G(y)\overline{F}(y)^2dF(y)dG(x) \\ &+ 4(n-2)\{\frac{4}{3}\int_{-\infty}^{\infty}F(x)^3G(x)dG(x) - 2\int_{-\infty}^{\infty}F(x)G(x)dG(x) \\ &- 2\int_{-\infty}^{\infty}F(x)^2G(x)^2dG(x) + 3\int_{-\infty}^{\infty}G(x)^2F(x)dG(x) \\ &+ \int_{-\infty}^{\infty}F(x)^2dG(x) - \int_{-\infty}^{\infty}F(x)^3dG(x)\}, \text{ where } F(x) = 1-\overline{F}(x) \\ &\qquad \text{and } G(x) = 1-\overline{G}(x). \end{aligned} \tag{2.3}$$

The asymptotic normality of $\hat{\tau}$ follows by the results of Hoeffding (1948). Hence, a test of independence versus dependence rejects if $|\hat{\tau}/\sqrt{V(\hat{\tau})}|$ is greater than the appropriate percentage point of a standard normal random variable. A test of independence versus positive dependence rejects if $\hat{\tau}/\sqrt{V(\hat{\tau})}$ is too large.

The variance of $\hat{\tau}$ (2.3) can be expressed explicitly in several cases.

Case 1. $\bar{F}(x) = \bar{G}(x)$. In this case (2.3) reduces to

$$V(\hat{\tau}) = \frac{4n+7}{30n(n-1)} . \tag{2.4}$$

Case 2. (Lehmann structure) $\bar{F}(x) = \bar{G}(x)^{\alpha}$. Here (2.3) reduces to

$$n(n-1)V(\hat{\tau}) = \frac{8\alpha[35\alpha + n(9\alpha^2+2\alpha+9)]}{3(3\alpha+1)(3+\alpha)(2\alpha+3)(3\alpha+2)} . \tag{2.5}$$

Case 3. (X, Y exponential), $\bar{F}(X) = e^{-\lambda x}$, $\bar{G}(y) = e^{-\theta y}$, then (2.3) reduces to

$$n(n-1)V(\hat{\tau}) = \frac{8\lambda\theta[35\lambda\theta + n(9\lambda^2+2\lambda\theta+9\theta^2)]}{3(3\lambda+\theta)(\lambda+3\theta)(2\lambda+3\theta)(3\lambda+2\theta)} . \tag{2.6}$$

3. SIMULATION STUDY

A simulation study was conducted on the AMDAHL 470 computer at The Ohio State University. The study was performed by generating 1,000 samples of $n = 20$ or 40 series systems with exponentially distributed component life times, $\bar{F}(x) = e^{-x}$, and $\bar{G}(y) = e^{-\theta y}$, $\theta = 1., 2.$ The value of Kendall's $\tau = 0, .3, .6, .9$ was obtained by using the procedure of Johnson and Tenebein (1981).

Table 1 summarizes the results of this study. Emperical powers times 1000 of the modified test of H_0: independence versus H_1: $\tau > 0$ were computed by comparing $\hat{\tau}/\sqrt{V(\hat{\tau})}$ to the α^{th} upper percentage point of a standard normal. The resulting emperical powers (and their standard errors in ( ) along with the estimated expectation of $\hat{\tau}$ (and its standard error in ( )) are reported.

TABLE 1. SIMULATION STUDY OF $\hat{\tau}$

θ	n	τ	α = .1	α = .05	α = .025	α = .01	$E(\hat{\tau})$
1	20	0.0	97(9)	44(7)	17(4)	7(3)	-.0039(.0027)
1	40	0.0	84(9)	42(6)	22(4)	7(3)	-.0027(.0034)
1	20	0.3	356(15)	241(14)	168(12)	86(9)	.0794(.0029)
1	40	0.3	514(15)	376(15)	275(14)	177(12)	.0766(.0021)
1	20	0.6	726(14)	593(16)	461(16)	317(15)	.1596(.0028)
1	40	0.6	902(9)	831(12)	734(14)	591(15)	.1596(.0020)
1	20	0.9	880(10)	822(12)	733(14)	612(15)	.2251(.0029)
1	40	0.9	998(3)	970(5)	951(7)	899(9)	.2214(.0020)
2	20	0.0	92(9)	46(7)	20(4)	4(2)	-.0028(.0027)
2	40	0.0	101(9)	55(7)	23(5)	9(3)	.0009(.0019)
2	20	0.3	426(16)	319(15)	233(13)	139(11)	.0952(.0030)
2	40	0.3	599(15)	463(16)	338(15)	222(13)	.0891(.0021)
2	20	0.6	802(13)	711(14)	605(16)	478(16)	.1923(.0029)
2	40	0.6	964(6)	929(8)	875(10)	784(13)	.1910(.0020)
2	20	0.9	998(1)	997(2)	986(4)	950(7)	.3154(.0021)
2	40	0.9	1000(0)	1000(0)	1000(0)	1000(0)	.3169(.0014)

Note: Numbers reported are the estimates (standard error of the estimates).

Several conclusions can be drawn from this study. First, the normal approximation to the null distribution of $\hat{\tau}$ appears to hold quite well for relatively small sample sizes like 20. Secondly, $E(\hat{\tau})$ is positive for $\tau > .3$ and in fact we conjecture this is true for $\tau > 0$. However, as expected $\hat{\tau}$ is underestimating τ by a substantial margin and as such is not a good point estimator. Thirdly, $\hat{\tau}$ has reasonably good power when τ is large. Lastly, the power at fixed n, and τ is higher when the components are not identical. This is reasonable since in this case our test is based on T = minimum of (X, Y) and not on the distribution of (T, I = 1) and (T, I = 2) the crude probabilities.

4. CONCLUDING REMARKS

This article presents a technique for testing the independence of component lifetimes based on data collected when the components are linked in series. The problem of nonidentifiability which prohibits testing, nonparametrically, this hypothesis of independence is circumvented, in most cases, by assuming the marginal component life distributions are known. In general, this knowledge of the component distributions is not sufficient to resolve the identifiability dilemma. That is, there is a small class of distributions, mathematically contrived, which have the given set of marginal distributions and lead to the same observable information as a set of independent component lifetimes. Our test will have no power to detect these departures from independence. However, for a given set of marginals there is a large class of joint distributions where knowledge of the marginals and the system information is sufficient to determine independence. This class contains the standard bivariate distributions of interest in modeling series systems. For distributions in this class our simulation study shows that this test performs well for relatively small sample sizes.

5. ACKNOWLEDGMENTS

The author would like to thank Mr. Sukhoon Lee for his assistance in programming the simulation study. Also Professors M. L. Moeschberger of Ohio State University and P. Laud of Northern Illinois for their helpful comments. This research was supported by Contract AFOSR-82-0307 for the Air Force Office of Scientific Research.

REFERENCES

Basu, A. P. and Klein, J. P. (1982). "Some Recent Results on Competing Risks Theory" in Survival Analysis, eds. J. Crowley and R. Johnson, Hayward, CA.: Institute of Mathematical Statistics, 216-229.

David, H. A. and Moeschberger, M. L. (1978). The Theory of Competing Risks, New York: MacMillan Publishing Co., Inc.

Easterling, R. G. and Prairie, R. R. "Combining Component and System Information", Technometrics, 13, 271-280.

Hoeffding, W. (1948). "A Class of Statistics with Asymptotically Normal Distribution", Annals of Mathematical Statistics, 19, 293-325.

Johnson, M. and Tenenhein, A. (1981). "A Bivariate Distribution Family with Specified Marginals", Journal of the American Statistical Association, 76, 198-201.

Kendall, M. (1938). "A New Measure of Rank Correlation", Biometrika, 30, 81-93.

Klein, J. P. and Moeschberger, M. L. (1983). "Consequences of Assuming Independence in a Bivariate Exponential Series System", Technical Report No. 281, Department of Statistics, The Ohio State University.

Mastran, D. V. (1976). "Incorporating Component and System Data into the Same Assessment: A Bayesian Approach", Operation Research Society Journal, 24, 491-499.

Miller, D. R. (1977). "A Note on Independence of Multivariate Lifetimes in Competing Risks", Annals of Statistics, 5, 576-579.

Miyamura, T. (1982). "Estimating Component Failure Rates from Combined Component and System Data: Exponentially Distributed Component Lifetimes", Technometrics, 24, 313-318.

Moeschberger, M. L. and Klein, J. P. (1982). "Robustness to Departures from Independence in Exponential Series Systems", Submitted.

Tsiatis, A. (1975). "A Nonidentifiability Aspect of the Problem of Competing Risks", Proceedings of the National Academy of Sciences, 72, 20-22.

APPENDIX

To show $E(\hat{\tau}) = 0$ and $V(\hat{\tau}) = (2.3)$ under H_o consider the pairs (T_1, I_1), (T_2, I_2) and (T_3, I_3). Let $A_1 = \{T_1 < T_2, I_1 = I_2 = 1\}$, $A_2 = \{T_1 < T_2, I_1 = 1, I_2 = 0\}$, $A_3 = \{T_1 < T_2, I_1 = 0, I_2 = 0\}$ and $A_4 = \{T_1 < T_2, I_1 = 0, I_2 = 1\}$. In terms of the unobservable component lifetimes, (X_i, Y_i), $A_1 = \{X_1 < X_2, X_1 < Y_1, X_2 < Y_2\}$, $A_2 = \{X_1 < Y_2, X_1 < Y_1, Y_2 < X_2\}$, $A_3 = \{Y_1 < Y_2, Y_1 < X_1, Y_2 < X_2\}$, and $A_3 = \{Y_1 < X_2, Y_1 < X_1, X_2 < Y_2\}$. Under the null hypothesis of independence T_1 is equally likely to be either smaller or larger than T_2 so

$$\frac{1}{2} E(\phi(T_1,I_1,T_2,I_2)) = \int_{A_1} \frac{\bar{G}(x_1)-\bar{G}(x_2)}{\bar{G}(x_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$

$$+ \int_{A_2} \frac{\bar{G}(x_1)-2\bar{G}(y_2)}{\bar{G}(x_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$

$$+ \int_{A_3} \frac{\bar{F}(y_1)-\bar{F}(y_2)}{\bar{F}(y_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2) \qquad \text{(A.1)}$$

$$+ \int_{A_4} \frac{\bar{F}(y_1)-2\bar{F}(x_2)}{\bar{F}(y_1)} dF(x_1)dF(x_2)dG(y_1)dG(y_2).$$

$$= J_1 + J_2 + J_3 + J_4 \quad \text{(say)}.$$

Now, consider

$$J_1 + J_2 = \int_{-\infty}^{\infty} \{\int_x^{\infty} [\bar{G}(x)-\bar{G}(y)]\bar{G}(y)dF(y) + \int_x^{\infty} [\bar{G}(x)-2\bar{G}(y)]\bar{F}(y)dG(y)\}dF(x). \qquad \text{(A.2)}$$

Integrating the first inter integral in (A.2) by parts yields the negative of the second inter integral so that $J_1 + J_2 = 0$. Similar computations show that $J_3 + J_4 = 0$. Thus $E(\phi(T_1, I_1, T_2, I_2))$ and hence $E(\hat{\tau})$ are both 0.

Now

$$\frac{1}{2}E(\phi^2(T_1,I_1,T_2,I_2)) = \int_{A_1}\left[\frac{\bar{G}(x_1)-\bar{G}(x_2)}{\bar{G}(x_1)}\right]^2 dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$

$$+\int_{A_2}\left[\frac{\bar{G}(x_1)-2\bar{G}(y_2)}{\bar{G}(x_1)}\right]^2 dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$

$$+\int_{A_3}\left[\frac{\bar{F}(y_1)-\bar{F}(y_2)}{\bar{F}(y_1)}\right]^2 dF(x_1)dF(x_2)dG(y_1)dG(y_2) \qquad (A.3)$$

$$+\int_{A_4}\left[\frac{\bar{F}(y_1)-2\bar{F}(x_2)}{\bar{F}(y_2)}\right] dF(x_1)dF(x_2)dG(y_1)dG(y_2)$$

$$= J_1' + J_2' + J_3' + J_4' \quad \text{(say)}.$$

After a little simplification we have

$$J_1' + J_2' = \int_{-\infty}^{\infty}\frac{1}{\bar{G}(x_1)}\left\{\int_x^{\infty}[\bar{G}(x)-\bar{G}(y)]^2\bar{G}(y)dF(y) + \int_x^{\infty}[\bar{G}(x)-2\bar{G}(y)]^2\bar{F}(y)dG(y)\right\}dF(x)$$

$$= \int_{-\infty}^{\infty}\frac{1}{\bar{G}(x)}\left\{\int_x^{\infty}\bar{G}^2(y)\bar{F}(y)dG(y)\right\}dF(x) \qquad (A.4)$$

after integrating the first term by parts. Writing $\bar{F}(y) = 1 - F(y)$ it follows that

$$J_1' + J_2' = \frac{1}{3}\int_{-\infty}^{\infty}\bar{G}(x)^2 dF(x) - \int_{-\infty}^{\infty}\frac{1}{\bar{G}(x)}\int_x^{\infty}\bar{G}(y)^2F(y)dG(y)dF(x). \qquad (A.5)$$

Similarly $J_3' + J_4'$ is of the form (A.5) with the rolls of F and G reversed. Hence

$$E(\phi^2(T_1,I_1,T_2,I_2)) = \frac{2}{3}\int_{-\infty}^{\infty}\bar{G}(x)^2dF(x) + \frac{2}{3}\int_{-\infty}^{\infty}\bar{F}(x)^2dG(x)$$

$$- 2\int_{-\infty}^{\infty}\frac{1}{\bar{G}(x)}\int_x^{\infty}F(y)\bar{G}(y)^2dG(y)dF(x) - 2\int_{-\infty}^{\infty}\frac{1}{\bar{F}(x)}\int_x^{\infty}G(y)\bar{F}(y)^2dF(y)dG(x). \qquad (A.6)$$

Now, under independence,

$E[\phi(T_1,I_1,T_2,I_2)\phi(T_1,I_1,T_3,I_3)] =$

$$\int_{\{x_1<x_1,x_2,x_1<y_1,x_2<y_2,x_3<y_3\}} \frac{[\bar{G}(x_1)-\bar{G}(x_2)]}{\bar{G}(x_1)} \frac{[\bar{G}(x_1)-\bar{G}(x_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.7}$$

$$+ \int_{\{x_2,x_3<x_1,x_1<y_1,x_2<y_2,x_3<y_3\}} \frac{[\bar{G}(x_2)-\bar{G}(x_1)]}{\bar{G}(x_2)} \frac{[\bar{G}(x_3)-\bar{G}(x_1)]}{\bar{G}(x_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.8}$$

$$+ 2\int_{\{x_2<x_1<x_3,x_1<y_1,x_2<y_2,x_3<y_3\}} \frac{[\bar{G}(x_2)-\bar{G}(x_1)]}{\bar{G}(x_2)} \frac{[\bar{G}(x_1)-\bar{G}(x_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.9}$$

$$+ \int_{\{y_1<y_2,y_3,y_1<x_1,y_2<x_2,y_3<x_3\}} \frac{[\bar{F}(y_1)-\bar{F}(y_2)]}{\bar{F}(y_1)} \frac{[\bar{F}(y_1)-\bar{F}(y_3)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.10}$$

$$+ \int_{\{y_2,y_3<y_1,y_1<x_1,y_2<x_2,y_3<x_3\}} \frac{[\bar{F}(y_2)-\bar{F}(y_1)]}{\bar{F}(y_2)} \frac{[\bar{F}(y_3)-\bar{F}(y_1)]}{\bar{F}(y_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.11}$$

$$+ 2\int_{\{y_2<y_1<y_3,y_1<x_1,y_2<x_2,y_3<x_3\}} \frac{[\bar{F}(y_2)-\bar{F}(y_1)]}{\bar{F}(y_2)} \frac{[\bar{F}(y_1)-\bar{F}(y_3)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.12}$$

$$+ \int_{\{x_1<y_2,y_3,x_1<y_1,y_2<x_2,y_3<x_3\}} \frac{[\bar{G}(x_1)-2\bar{G}(y_2)]}{\bar{G}(x_1)} \frac{[\bar{G}(x_1)-\bar{G}(y_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.13}$$

$$+ \int_{\{y_2,y_3<x_1,\,y_2<x_2,\,y_3<x_3,\,x_1<y_1\}} \frac{[\bar{F}(y_2)-2\bar{F}(x_1)]}{\bar{F}(y_2)} \frac{[\bar{F}(y_3)-2\bar{F}(x_1)]}{\bar{F}(y_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.14}$$

$$+ 2\int_{\{y_2<x_1<y_3,\,y_2<x_2,\,x_1<y_2,\,y_3<x_3\}} \frac{[\bar{F}(y_2)-2\bar{F}(x_1)]}{\bar{F}(y_2)} \frac{[\bar{G}(x_1)-2\bar{G}(y_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.15}$$

$$+ \int_{\{y_1<x_2,x_3,\,y_1<x_1,\,x_2<y_2,\,x_3<y_3\}} \frac{[\bar{F}(y_1)-2\bar{F}(x_2)]}{\bar{F}(y_1)} \frac{[\bar{F}(y_1)-2\bar{F}(x_3)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.16}$$

$$+ \int_{\{x_2,x_3<y_1,\,y_1<x_1,\,x_2<y_2,\,x_3<y_3\}} \frac{[\bar{G}(x_2)-2\bar{G}(y_1)]}{\bar{G}(x_2)} \frac{[\bar{G}(x_3)-2\bar{G}(y_1)]}{\bar{G}(x_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.17}$$

$$+ 2\int_{\{x_2<y_1<x_3,\,x_2<y_2,\,y_1<x_1,\,x_3<y_3\}} \frac{[\bar{G}(x_2)-2\bar{G}(y_1)]}{\bar{G}(x_2)} \frac{[\bar{F}(y_1)-2\bar{F}(x_3)]}{\bar{F}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.18}$$

$$+ 2\int_{\{x_1<x_2,y_3,\,x_1<y_1,\,x_2<y_2,\,y_3<x_3\}} \frac{[\bar{G}(x_1)-\bar{G}(x_2)]}{\bar{G}(x_1)} \frac{[\bar{G}(x_1)-2\bar{G}(y_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.19}$$

$$+ 2\int_{\{x_2,y_3<x_1,\,x_1<y_1,\,x_2<y_2,\,y_3<x_3\}} \frac{[\bar{G}(x_2)-\bar{G}(x_1)]}{\bar{G}(x_2)} \frac{[\bar{F}(y_3)-2\bar{F}(x_1)]}{\bar{F}(y_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.20}$$

$$+ 2\int_{\{x_2<x_1<y_3,\,x_2<y_2,\,x_1<y_1,\,y_3<x_3\}} \frac{[\bar{G}(x_2)-\bar{G}(x_1)]}{\bar{G}(x_2)} \frac{[\bar{G}(x_1)-2\bar{G}(y_3)]}{\bar{G}(x_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.21}$$

$$+ 2 \int_{\{y_3<x_1<x_2,y_3<x_3,x_1<y_1,x_2<y_2\}} \frac{[\bar{G}(x_1)-\bar{G}(x_2)]}{\bar{G}(x_1)} \frac{[\bar{F}(y_3)-2\bar{F}(x_1)]}{\bar{F}(y_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.22}$$

$$+ 2 \int_{\{y_1<x_2,y_3,y_1<x_1,x_2<y_2,y_3<x_3\}} \frac{[\bar{F}(y_1)-2\bar{F}(x_2)]}{\bar{F}(y_1)} \frac{[\bar{F}(y_1)-\bar{F}(y_3)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.23}$$

$$+ 2 \int_{\{x_2,y_3<y_1,x_2<y_2,y_3<x_3,y_1<x_1\}} \frac{[\bar{G}(x_2)-2\bar{G}(y_1)]}{\bar{G}(x_2)} \frac{[\bar{F}(y_3)-\bar{F}(y_1)]}{\bar{F}(y_3)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.24}$$

$$+ 2 \int_{\{x_2<y_1<y_3,x_2<y_2,y_1<x_1,y_3<x_3\}} \frac{[\bar{G}(x_2)-2\bar{G}(y_1)]}{\bar{G}(x_2)} \frac{[\bar{F}(y_1)-\bar{F}(y_3)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i) \tag{A.25}$$

$$+ 2 \int_{\{y_3<y_1<x_2,y_3<x_3,y_1<x_1,x_2<y_2\}} \frac{[\bar{F}(y_3)-\bar{F}(y_1)]}{\bar{F}(y_3)} \frac{[\bar{F}(y_1)-2\bar{F}(x_2)]}{\bar{F}(y_1)} \prod_{i=1}^{3} dF(x_i)dG(y_i). \tag{A.26}$$

Now, by combining like terms and a integation by parts one can verify that

(A.7) + (A.13) + (A.19) = 0, (A.10) + (A.16) + (A.23) = 0,

(A.9) + (A.15) + (A.21) + (A.22) = 0,

(A.12) + (A.18) + (A.25) + (A.26) = 0,

$$(A.8) + (A.14) + (A.20) = \int_{-\infty}^{\infty} \bar{G}(x)G^2(x)[2F(x)-1]^2 dF(x) \tag{A.27}$$

$$(A.11) + (A.17) + (A.24) = \int_{-\infty}^{\infty} \bar{F}(x)F^2(x)[2G(x)-1]^2 dG(x). \tag{A.28}$$

Integrating (A.27) by parts and combining the result with (A.28) we have

$$E(\phi(T_1,I_1,T_2,I_2)\phi(T_1,I_1,T_3,I_3)) =$$

$$\frac{4}{3}\int_{-\infty}^{\infty} F(x)^3G(x)dG(x) - 2\int_{-\infty}^{\infty} F(x)G(x)dG(x) - 2\int_{-\infty}^{\infty} F(x)^2G(x)^2dG(x)$$
$$+ 3\int_{-\infty}^{\infty} G(x)^2F(x)dG(x) + \int_{-\infty}^{\infty} F(x)^2dG(x) - \int_{-\infty}^{\infty} F(x)^3dG(x). \quad (A.29)$$

To find $V(\hat{\tau}) = E(\sum\sum_{i<j} \phi(T_i,I_i,T_j,I_j)/\binom{n}{2})^2$

$$= [2E(\phi(T_1,I_1,T_2,I_2)^2) + 4(n-2)E(\phi(T_1,I_1,T_2,I_2)\phi(T_1,I_1,T_3,I_3))]/n(n-1)$$

which is equal to (2.3) after making the substitutions (A.6) and (A.29).

DATE
ILME