

AD-A037 969

FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS
TRANSFORMATION YIELDING RELIABILITY MODELS BASED ON INDEPENDENT--ETC(U)
JAN 77 N LANGBERG, A J QUINZI, F PROSCHAN

F/G 12/1

AF-AFOSR-2581-74

UNCLASSIFIED

FSU-STATISTICS-M399

AFOSR-TR-77-0254

NL

1 OF 1
AD
A037969



END

DATE
FILMED

5-77

19 REPORT DOCUMENTATION PAGE

**READ INSTRUCTIONS
BEFORE COMPLETING FORM**

AD A 037969 (6)

Approved for public release; distribution unlimited.

DDC
RECEIVED
APR 8 1977
RECEIVED

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

AJ 1's.
DOC FILE COPY

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

→ This is an expository paper presenting various ways of transforming dependent models into independent ones and displaying applications in variety of contexts including reliability modelling, life testing, and nonparametric estimation in the study of competing risks.

400 27%

~~UNCLASSIFIED~~

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR - TR - 77 - 0254

ADDITIONAL	
DDC	Write Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNCLASSIFIED	<input type="checkbox"/>
JUSTIFICATION	
DISTRIBUTION/AVAILABILITY CODES	
DIST.	AVAIL. and/or SPECIAL
A	

Transformations Yielding Reliability Models Based
on Independent Random Variables: A Survey

by

N. Langberg¹, F. Proschan¹, and A. J. Quinzi²

FSU Statistics Report M399
AFOSR Technical Report No. 69

January, 1977
The Florida State University
Department of Statistics
Tallahassee, Florida

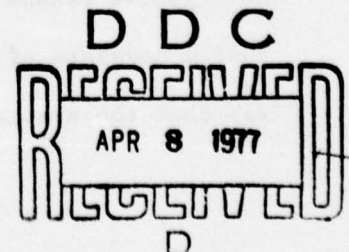
Approved for public release;
distribution unlimited.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

¹Florida State University

²Florida State University and Temple University

Research supported by the Air Force Office of Scientific Research under
AFOSR Grant 74-2581 .



Abstract. This is an expository paper presenting various ways of transforming dependent models into independent ones and displaying applications in a variety of contexts including reliability modelling, life testing, and nonparametric estimation in the study of competing risks.

1. Introduction.

The central theme of this survey is the transformation of dependent models into independent ones. By "dependent (independent) models" we mean multivariate stochastic models whose joint probability distribution are distributions of dependent (independent) random variables. Each of the transformations discussed here can be used to convert the original dependent model into an independent model which is equivalent (in a specified sense) to the original model. It is the purpose of this paper to present:

- (a) some key theorems upon which such transformations are based, and
- (b) a variety of applications in reliability and biometry.

We do not give formal proofs of the results presented; these may be found in the original papers cited. Rather, we motivate the key ideas by examining important special cases and several illustrative examples.

2. Distributions with exponential and proportional hazard minima.

In this section and the next we describe methods for converting dependent models into independent ones based upon the assumption that the joint distribution of the random variables in the original (dependent) model belongs to a specified family of distributions.

We begin with some terminology and notation. A life length T is a nonnegative random variable such that $\lim_{t \rightarrow \infty} P(T > t) = 0$. Suppose that a system consists of n components with random life lengths T_1, \dots, T_n . We say that the system is a series (parallel) system if the failure of the system

coincides with the earliest (latest) component failure. Thus, the life length of the corresponding series (parallel) system is given by $\min(T_i, 1 \leq i \leq n)$ [$\max(T_i, 1 \leq i \leq n)$]. Series and parallel systems are examples of more general systems in reliability known as coherent systems [see Birnbaum, Esary, and Saunders (1961) or Barlow and Proschan (1975)].

A random vector (T_1, \dots, T_n) has exponential minima if $\min(T_i, i \in I)$ is exponentially distributed for every nonempty subset I of $\{1, \dots, n\}$. In reliability terms, a random vector has exponential minima if the life length of every series subsystem (i.e., every series system which may be formed by using a subset of the n components) is exponentially distributed. In particular, individual components (series systems of size one) have exponential life lengths. The $(n\text{-dimensional})$ random vectors \underline{T} and \underline{U} are marginally equivalent in minima ($\underline{T} \stackrel{m}{=} \underline{U}$, in symbols) if $\min(T_i, i \in I)$ and $\min(U_i, i \in I)$ have the same distribution for each nonempty $I \subset \{1, \dots, n\}$.

A particularly important multivariate distribution with exponential minima is the multivariate exponential (MVE) of Marshall and Olkin (1967). The classic paper of Marshall and Olkin (1967) and the model derived therein have prompted numerous investigations [see the annotated bibliography of Kotz (1974)]. The following characterization of the MVE is a particularly useful one.

Theorem 2.1. A random vector (U_1, \dots, U_n) has the $(n\text{-dimensional})$ MVE distribution if and only if there exists a collection $\{H_J, J \in J\}$, $UJ = \{1, \dots, n\}$, of independent exponential random variables such that $U_i = \min(H_J, J \in J, i \in J)$, $i = 1, \dots, n$.

Theorem 2.1 is an immediate extension of Theorem 3.2 of Marshall and Olkin (1967). To simplify notation we shall adopt the following conventions throughout the remainder of this paper. Let I denote the collection of all nonempty subsets of $\{1, \dots, n\}$. Whenever an element $\{i_1, \dots, i_m\}$

of I appears as a subscript, as, for example, in $H_{\{i_1, \dots, i_m\}}$, we shall often write instead $H_{i_1 \dots i_m}$. We say that two random variables X and Y are stochastically equal and write $X \stackrel{st}{=} Y$, if X and Y have the same probability distribution. Unless otherwise indicated, all random vectors are assumed to be n -dimensional.

Esary and Marshall (1974) prove the following:

Theorem 2.2. Suppose that a random vector \underline{T} has exponential minima. Then there exists a random vector \underline{U} with the MVE distribution of Marshall and Olkin such that $\underline{T} \stackrel{m}{=} \underline{U}$.

Theorem 2.2 can be used to obtain a consistent estimator for system reliability as the following example illustrates:

Example 2.3. Suppose that an estimate of system reliability for an arbitrary coherent system of n components is desired prior to manufacture of the system. Suppose also that the only available failure data, however, is for n-component parallel systems whose component life lengths have the same joint distribution as those of the given system. If component life lengths have exponential minima, then by Theorem 2.2 there is a random vector \underline{U} with the MVE distribution such that $\underline{T} \stackrel{m}{=} \underline{U}$. Consistent estimators for the parameters of the MVE, given the failure data from parallel systems as above, have been obtained by Proschan and Sullo (1976). Since the reliability of the system can be expressed as a continuous function of survival probabilities $P[\min(T_i, i \in I) > t]$, $I \in I$ [see Esary and Marshall (1970)], we can replace $P[\min(T_i, i \in I) > t]$ by an estimator for $P[\min(U_i, i \in I) > t]$ given by Proschan and Sullo (1976) and thus obtain a consistent estimator for system reliability.

In view of Theorem 2.1, we can state Theorem 2.2 in the following equivalent form.

Theorem 2.3. Suppose that a random vector \underline{T} has exponential minima. Then there exists a collection $\{H_J, J \in J\}$, $UJ = \{1, \dots, n\}$, of independent exponential random variables such that for each $I \in I$,

$$\min(T_i, i \in I) \stackrel{\text{st}}{=} \min(H_J, J \in J, J \cap I \neq \emptyset).$$

Consider, for example, a series system whose component life lengths T_1, \dots, T_n have exponential minima and are mutually dependent. If we view the independent random variables $\{H_J\}$ of Theorem 2.3 as the component life lengths in a new series system, then, in effect, Theorem 2.3 allows us to transform a dependent model into an independent one, while preserving the life distribution of the original system. Under different conditions we shall see in Section 4 how to transform a dependent model into an independent one, while preserving not only system life length, but also the probabilities of occurrence of certain "failure patterns". We shall also see that such transformations from dependent to independent models are not only of interest in their own right, but also have important statistical applications to the theory of competing risks and to statistical life testing, in general.

Esary and Marshall (1974) establish existence only in their proof of Theorem 2.3 by using the special nature of coherent systems in reliability theory. The proof of Theorem 2.3 given by Langberg, Proschan, and Quinzi (1977a) is considerably more elementary and specifies explicitly the distributions of the independent random variables $\{H_J\}$ as follows. Suppose that $P[\min(T_i, i \in I) > t] = \exp(-\mu_I t)$, $I \in I$, for some collection $\{\mu_I, I \in I\}$ of positive constants. Then the random variable H_J in Theorem 2.3 is exponentially distributed with parameter λ_J given by

$$\lambda_J = (-1)^{\#(J)-1} (\mu_{1 \dots n} - \sum_{i \in J} \mu_{\{i\}} + \sum_{\substack{i_1, i_2 \in J \\ i_1 < i_2}} \mu_{\{i_1, i_2\}} - \dots + (-1)^{\#(J)} \mu_J), \quad (2.1)$$

where $\#(J)$ is the cardinality of J and \bar{J} is the complement of J in $\{1, \dots, n\}$. Formula (2.1) provides an explicit solution for the parameters $\{\lambda_J\}$ in terms of the known constants $\{\mu_I\}$. Formula (2.1) also indicates ways of testing the validity of the assumption of exponential minima as the following examples illustrate.

First, suppose it is known a priori that, due to the structure of a particular system, it is impossible for the components in some subset J (generally, some collection $\{J_1, \dots, J_m\}$ of subsets) to fail simultaneously. This is equivalent to assuming that the corresponding parameter $\lambda_J = 0$. If the corresponding linear combination of (known) constants $\{\mu_I\}$ given by (2.1) does not yield $\lambda_J = 0$, then the assumption of exponential minima must be wrong. Similarly, if some combination of the $\{\mu_I\}$ yields a λ_J which is negative, then the assumption of exponential minima is likewise incorrect.

More generally, formula (2.1) indicates a heuristic method for testing the statistical hypothesis of exponential minima. For example, consider a two-component system with component life lengths T_1 and T_2 . If $P[\min(T_i, i \in I) > t] = \exp(-\mu_I t)$, $I \in \mathcal{I}$, then by (2.1),

$$\lambda_1 = \mu_{12} - \mu_2 \geq 0$$

$$\lambda_2 = \mu_{12} - \mu_1 \geq 0$$

$$\lambda_{12} = \mu_1 + \mu_2 - \mu_{12} \geq 0.$$

Consequently, we would expect that estimates for the μ_I 's, together with an allowance for random error, would satisfy a similar set of inequalities. If not, we would tend to doubt the hypothesis of exponential minima.

Employing the same technique of proof used to prove Theorem 2.3, Langberg, Proschan, and Quinzi (1977a) [hereafter referred to as LPQ (1977a)] obtain a generalization of Theorem 2.3. First we introduce some terminology. The

hazard function R associated with the distribution function F of a nonnegative random variable is the function $R(t) = -\log[1 - F(t)]$, $t \geq 0$. The (multivariate distribution of the) nonnegative random vector \underline{T} has proportional hazard minima if there exists a collection $\{\mu_I, I \in I\}$ of positive constants such that $P[\min(T_i, i \in I) > t] = \exp[-\mu_I R(t)]$, $I \in I$, where $R(\cdot)$ is a hazard function, i.e., a nonnegative, nondecreasing function satisfying $R(0) = 0$ and $R(\infty) = \infty$. LPQ (1977a) prove the following:

Theorem 2.4. Suppose that a random vector \underline{T} has proportional hazard minima with hazard function R . Suppose further that R is continuous at $t_0 = \sup\{t: R(t) = 0\}$. Then there exists a collection $\{H_J, J \in J\}$, $\cup J = \{1, \dots, n\}$, of independent random variables with hazard functions proportional to $R(\cdot)$ such that for each $I \in I$,

$$\min(T_i, i \in I) \stackrel{\text{st}}{=} \min(H_J, J \in J, J \cap I \neq \emptyset).$$

Furthermore, the constants of proportionality $\{\lambda_J, J \in J\}$ are given by (2.1).

Remark 2.5. Note that in the special case $R(t) = t$, $t \geq 0$, the conclusion of Theorem 2.4 holds by Theorem 2.3.

The assumption of proportional hazard minima holds, for example, when the random vector \underline{T} has minima with the Weibull distribution having a fixed scale parameter.

3. Additive families of distributions.

Again using the same technique of proof as in the proof of Theorem 2.3, LPQ (1977a) obtain a result for additive families of distributions which is analogous to Theorem 2.3 except that "sum" plays the role of "minimum".

Let $F = \{F_\theta, \theta \in \Theta\}$ be a family of distributions parameterized by θ , and let g be a binary operation on the set of real numbers. The collection

F is said to form an additive family with respect to g if for every $\theta_1, \theta_2 \in \Theta$: X_1 and X_2 are independent random variables with respective distributions F_{θ_1} and F_{θ_2} implies that $g(X_1, X_2)$ has distribution $F_{\theta_1 + \theta_2}$. In Theorem 2.3, the family of interest is the additive family of exponential distributions, $F_{\theta}(x) = 1 - \exp(-\theta x)$, with respect to $g(x_1, x_2) = \min(x_1, x_2)$. The following result applies to additive families $\{F_{\theta}, \theta \in \Theta\}$ with respect to $g(x_1, x_2) = x_1 + x_2$, where $\int x dF_{\theta}(x) = \theta$.

Theorem 3.1. For each $I \in I$, let the random variable T_I have distribution $F_{\mu_I} \in F = \{F_{\mu}, \mu \in M\}$, where F is an additive family of distributions with respect to $g(x_1, x_2) = x_1 + x_2$ satisfying $\int x dF_{\mu}(x) = \mu$. Then there exists a collection $\{S_J, J \in J\}$ of independent random variables such that the distribution of S_J belongs to F , $J \in J$, and for each $I \in I$,

$$T_I \stackrel{\text{st}}{=} \sum_{J \in J: J \cap I \neq \emptyset} S_J.$$

Furthermore, the mean λ_J of the random variable S_J is given by (2.1).

Examples of families satisfying the hypothesis of Theorem 3.1 are the Poisson family with mean μ and the gamma family with mean μ and unit scale parameter. A sample application in reliability follows.

Example 3.2. Suppose that an n -component system is exposed to shocks which are not necessarily fatal. For each $I \in I$, a shock of type I simultaneously affects all components exclusively in subset I . For example, the shock pattern for a two-component system might be exhibited as in Figure 3.1 below:

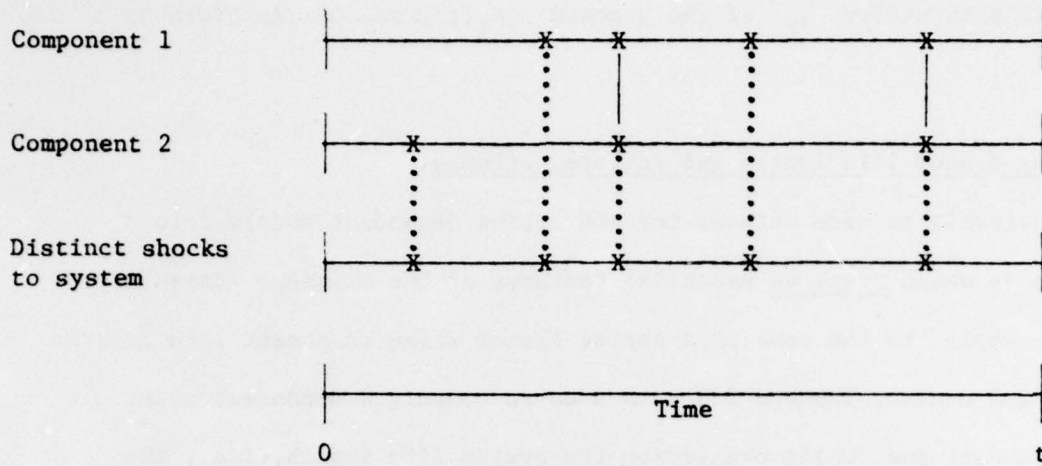


Figure 3.1.

Figure 3.1 indicates that a total of 5 distinct shocks occurred in the interval $[0, t]$: 2 shocks affected component 1 alone, 1 shock affected component 2 alone, and 2 shocks affected both components simultaneously. Let $N_I(t)$ be the number of shocks in the interval $[0, t]$ which are simultaneously received by the components exclusively in subset I . In Figure 3.1, $N_1(t) = 2$, $N_2(t) = 1$, $N_{12}(t) = 2$. Let $K_I(t)$ be the number of distinct (in time) shocks in the interval $[0, t]$ which are received by the components in subset I . In Figure 3.1, $K_1(t) = 4$, $K_2(t) = 3$, $K_{12}(t) = 5$. Note that, in general,

$$K_I(t) = \sum_{J \cap I \neq \emptyset} N_J(t),$$

so that the processes $\{K_I(t), t \geq 0\}$, $I \in I$, are generally dependent. If $\{K_I(t), t \geq 0\}$ is a Poisson process with intensity $\mu_I > 0$, $I \in I$, then we conclude from Theorem 3.1 that there exists a collection $\{\{N_J^*(t), t \geq 0\}, J \in J\}$ of independent Poisson processes such that for every $I \in I$,

$$K_I(t) \stackrel{st}{=} \sum_{J \in J: J \cap I \neq \emptyset} N_J^*(t).$$

Furthermore, the intensity λ_I of the process $\{N_I(t), t \geq 0\}$ is given by (2.1).

4. Preserving system life length and failure patterns.

It is desirable to have methods for converting dependent models into independent ones which preserve essential features of the original (dependent) model. For example, in the case of a series system whose component life lengths have exponential minima, Theorem 2.3 allows us to convert a dependent model into an independent one, while preserving the system life length, i.e., the minimum of the component life lengths. In this section we show how, under more general conditions, it is possible to convert a dependent model into an independent one, while preserving features of the original model in addition to system life length.

Consider an arbitrary series system of n components. In many practical applications we are able to observe:

- (1) the time at which the system fails, and
- (2) the identity of the component or set of components which fails.

Note that although we use the language of reliability theory (series system, component, etc.), the general model has application in a variety of contexts. For example, in population mortality studies, the data on each subject includes (1) the age at death and (2) the cause of death. Suppose that an individual dies due to one of n possible causes. An individual, in this context, can be viewed as an n -component series system who dies due to the occurrence of one or more of the n possible causes. As another example, suppose a personnel study is undertaken to determine the departure patterns of employees in a large company. The data on each employee might consist of (1) the length of stay, i.e., the time from arrival to termination and (2) the reason for termination. In general, one could imagine any model where observations include (1) the time

at which a particular event occurs and (2) the identity of the cause or set of causes (among a finite number) which results in the occurrence of the event. Moreover, one or more of the "causes" might be identified with the withdrawal of a unit from observation, resulting in censored or truncated data. For convenience we continue to employ the language applicable in many other situations.

In this section we show how it is possible to replace a series system of dependent components by a series system of independent components while simultaneously preserving

- (1)' the distribution of the time to system failure, and
- (2)' the probability of occurrence of each failure pattern.

By "failure pattern" we mean, in the case of a series system, the failure of a set of components whose simultaneous failure causes (i.e., coincides with) the failure of the system.

We begin with some terminology and notation. If \underline{T} is the vector of component life lengths in an n -component system, we say that failure pattern I occurs, and write $\xi(\underline{T}) = I$, if the simultaneous failures of the components exclusively in subset I coincide with the failure of the system. Let \underline{S} and \underline{T} be the vectors of component life lengths of two systems with respective life lengths S and T . We say that the systems are equivalent in life lengths and patterns ($\underline{S} \stackrel{LP}{=} \underline{T}$, in symbols) if

$$P(S > t, \xi(\underline{S}) = I) = P(T > t, \xi(\underline{T}) = I)$$

for every $t \geq 0$ and every $I \in I$.

Miller (1977) proves the following existence result:

Theorem 4.1. Let T_i be the life length of component i , $i = 1, \dots, n$, and let T be the life length of the corresponding series system. Assume that the functions

$$F(t, I) = P(T \leq t, \xi(T) = I), I \in I \quad (4.1)$$

have no common discontinuities and that $P(T_i = T_j) = 0$ for $i \neq j$. Then there exists a vector \underline{S} of independent random variables such that $\underline{T} \stackrel{LP}{=} \underline{S}$, and at least one of the S_i is a life length. The distributions of S_1, \dots, S_n are uniquely determined on $\{t: \bar{F}(t) > 0\}$, where $\bar{F}(t) = P(T > t)$, $t \geq 0$.

We can paraphrase Theorem 4.1 as follows. Under the given hypothesis, the original (dependent) system with life length $T = \min(T_i, 1 \leq i \leq n)$ can be replaced by a system with life length $S = \min(S_i, 1 \leq i \leq n)$, where the S_i 's are independent random variables in such a way that $\underline{S} \stackrel{LP}{=} \underline{T}$. Tsatis (1975) proves a similar result in the context of competing risk theory by assuming that the joint distribution of T_1, \dots, T_n has continuous partial derivatives. It is noteworthy that the nature of the dependence in the original model is unspecified in Theorem 4.1, i.e., the original components might be dependent in any way whatsoever. LPQ (1977b) show that the assumption of no common discontinuities in Theorem 4.1 is a necessary as well as a sufficient condition for the replacement of a dependent model by an independent one. Moreover, we provide explicit expressions for the appropriate distributions in the independent model.

Before presenting the LPQ (1977b) result, we motivate the theorem as follows. Let $T = \min(T_i, 1 \leq i \leq n)$, where T_1, \dots, T_n are the (dependent) component life lengths. Let $\bar{F}(t, I)$ be the joint probability that the system survives beyond time t and failure pattern I occurs. For example, if $n = 3$, $\bar{F}(t, \{1,2\}) = P(T > t, T_1 = T_2 < T_3)$ so that ties are possible. The problem as posed in LPQ (1977b) can be stated as follows. Given the vector \underline{T} of (dependent) life lengths, determine a random vector \underline{S} such that $\underline{S} \stackrel{LP}{=} \underline{T}$, where S_1, \dots, S_n are expressible in terms of independent random variables. By "expressible" we mean that the S_i are either themselves independent random variables or else can be expressed as functions of independent random variables.

The solution is found by letting S_1, \dots, S_n be the life lengths of components in a theoretical n -component series system, where the components are exposed to shocks according to the following shock model. Each component fails if it receives a shock. Independent sources of shock are present in the environment— one source for each $I \in \mathcal{I}$. A shock from source I simultaneously kills the components exclusively in subset I . Let H_I denote the time (measured from the origin) until a shock from source I occurs. Then $S_i = \min(H_I, I \in \mathcal{I}, i \in I)$, $1 \leq i \leq n$, and $S = H$, where $S = \min(S_i, 1 \leq i \leq n)$ and $H = \min(H_I, I \in \mathcal{I})$.

Define

$$\xi^*(\underline{H}) = \begin{cases} I, & \text{if } H_I < H_J \text{ for each } J \neq I \\ \emptyset & \text{otherwise.} \end{cases}$$

To allow for simultaneous failures among the components in the original system, we permit the dimension of the vector \underline{H} to be greater than or equal to that of \underline{T} . Generally, if \underline{T} has dimension n , then the vector \underline{H} of times until shock has dimension (at most) $2^n - 1$. The subscripts on the components of \underline{H} are understood to be ordered lexicographically. It follows that $P(S > t, \xi(\underline{S}) = I) = P(T > t, \xi(\underline{T}) = I)$ if and only if

$$P(H > t, \xi^*(\underline{H}) = I) = P(T > t, \xi(\underline{T}) = I) \quad (4.1)$$

for each $t \geq 0$ and each $I \in \mathcal{I}$. If (4.1) holds for every subset I of $\{1, \dots, n\}$, we write $\underline{H} \stackrel{LP}{=} \underline{T}$. The problem will be solved if we determine independent random variables $H_I, I \in \mathcal{I}$, such that (4.1) holds for every $t \geq 0$ and every subset I of $\{1, \dots, n\}$. LPQ (1977b) prove the following:

Theorem 4.2. Let $T = \min(T_i, 1 \leq i \leq n)$ be the life length of an n -component series system, where T_i is the life length of component i , $i = 1, \dots, n$. Define

$\bar{F}(t, I) = P(T > t, \xi(T) = I)$ and $F(t, I) = P(T \leq t, \xi(T) = I)$, $I \in I$. Let $\bar{F}(t) = P(T > t)$ and $\alpha(F) = \sup\{x: \bar{F}(x) > 0\}$. Then the following statements hold:

- (i) A necessary and sufficient condition for the existence of a set of independent random variables $\{H_I, I \in I\}$ which satisfy $H \stackrel{LP}{=} T$, where $H = \min(H_I, I \in I)$, is that the functions $F(\cdot, I)$, $I \in I$, have no common discontinuities in the interval $[0, \alpha(F))$.
- (ii) The random variables $\{H_I, I \in I\}$ in (i) have corresponding survival probabilities $\{\bar{G}_I(\cdot), I \in I\}$ which are uniquely determined on the interval $[0, \alpha(F))$ as follows:

$$\bar{G}_I(t) = \exp\left\{-\int_0^t [dF^C(\cdot, I)/\bar{F}]\right\} \cdot \prod_{a_j(I) \leq t} [\bar{F}(a_j(I))/\bar{F}(a_j(I)^-)], \quad 0 \leq t < \alpha(F), \quad (4.2)$$

where $F^C(\cdot, I)$ is the continuous part of $F(\cdot, I)$, $\{a_j(I)\}_{j=1}^\infty$ is the set of discontinuities of $F(\cdot, I)$, $I \in I$, and the product over an empty set is defined as unity.

The following generalization of Theorem 4.2 for arbitrary (not necessarily coherent) systems also holds:

Theorem 4.3. Let T_i denote the life length of component i , $i = 1, \dots, n$, in an arbitrary n -component system with life length T . Define $F(\cdot, I)$ and $\alpha(F)$ as in Theorem 4.2. Then (i) and (ii) of Theorem 4.2 hold.

Example 4.4. Suppose that the vector (T_1, T_2) has a bivariate distribution with survival probability:

$$\bar{F}(t_1, t_2) = (1 + t_1 + t_2)^{-1}, \quad t_1 \geq 0, \quad t_2 \geq 0.$$

Note that T_1 and T_2 are mutually dependent. If T_1 and T_2 are the component life lengths in a two-component series system, we may conclude from (4.2) of Theorem 4.2 that the original system is equivalent in life

length and failure patterns to a system involving independent times H_1 and H_2 until shock, where

$$\bar{G}_i(t) = P(H_i > t) = (1 + 2t)^{-1/2}, \quad i = 1, 2, t \geq 0.$$

5. Applications to the theory of competing risks, life testing, and censored data problems.

The reader will note that in every model thus far considered, we have obtained explicit expressions for the appropriate distributions in the independent model, whereas existence alone is proven in other approaches. Thus, our results are not only more general but also more readily applicable, especially when explicit solutions are called for. The probabilistic results obtained are, of course, of interest in their own right since they facilitate the analysis of various dependent models. However, a significant statistical payoff is also derived from our approach. In this section we show how our probabilistic solution to the conversion problem of Theorem 4.2 can be used to unify the nonparametric approach to estimation problems in competing risk theory, life testing, and certain incomplete data problems.

The theory of competing risks derives its name from the fact that, during a person's lifetime, he is exposed to several risks of death (various fatal diseases, accidents, etc.) which can be viewed as "competing" for his life. A series system of r dependent components with life length $T = \min(T_i, 1 \leq i \leq r)$ such that failure pattern I occurs [$\xi(T) = I$] becomes, in the terminology of competing risks, an individual with life length $T = \min(T_i, 1 \leq i \leq r)$ exposed to r dependent risks of death, where T_i is the age at death if risk i were the only risk present in the environment, $1 \leq i \leq r$, and ξ is the cause of death, i.e., the subset I of $\{1, \dots, r\}$ such that $T = T_i$ for each $i \in I$ and $T \neq T_i$ for each $i \notin I$. When death results from a single cause, then ξ is the index i for which $T = T_i$. [In an incomplete or censored data problem,

one of the random variables T_j represents the time at which an individual becomes "unobservable" for a reason other than death, while the remaining variables typically represent various causes of death.] The biomedical researcher is interested in making inferences about unobservable quantities (viz., the random variables T_1, \dots, T_r) by using data from observable quantities - in this case, the lifetime T and cause ξ . In showing how Theorem 4.2 may be applied, we focus on the following question: How can we estimate the marginal survival probabilities corresponding to a given risk (or combination of risks) operating alone without competition from the other risks? That is, how can we estimate the $2^r - 1$ survival probabilities (so-called "net probabilities") $\bar{M}_J(t) = P[\min(T_j, j \in J) > t], J \subset \{1, \dots, r\}$?

Throughout the remainder of this paper let $(T_{1i}, \dots, T_{ri}), i = 1, \dots, n$, represent a random sample of size n from the joint distribution of the nonnegative random variables T_1, \dots, T_r . To conform to the usual notation, we reserve 'n' for sample size. Thus, let I now denote the collection of all nonempty subsets of the set $\{1, \dots, r\}$ of risks. We adopt all of the previous notation subject to the substitution of 'r' for 'n'. For each distribution F , let $\bar{F} (\equiv 1 - F)$ be the corresponding survival function. For each $I \in I$, let $M_I(t) = P(T_I \leq t)$, where $T_I = \min(T_i, i \in I)$. In the competing risk model, only the following are observed:

$$T_{0i} = \min(T_{1i}, \dots, T_{ri}), 1 \leq i \leq n,$$

and

$$\xi_i, 1 \leq i \leq n,$$

where $\xi_i = I$ if and only if $T = T_i$ for each $i \in I$ and $T \neq T_i$ for each $i \notin I$.

Let $0 \equiv T'_{00} \leq T'_{01} \leq \dots \leq T'_{0n} \equiv T_{\max}$ denote the ordered values of the

observations T_{01}, \dots, T_{0n} .

Consider now the following assumptions:

- (A1) The risks (i.e., the random variables T_1, \dots, T_r) are mutually independent.
- (A2) No ties are possible, i.e., $P(T_i = T_j) = 0$ for each $i \neq j$, $i, j = 1, \dots, r$.
- (A3) The distributions of T_1, \dots, T_r have no common discontinuities.
- (A4) The random variables T_1, \dots, T_r have a joint distribution which is absolutely continuous.

Note that Assumptions (A2) and (A3) together imply that death results from a single cause.

Assuming (A1), (A2), and (A3), Peterson (1975) examines the following estimator for \bar{M}_J :

$$\hat{\bar{M}}_J(t) = \prod_R [(n - R)/(n - R + 1)], \quad (4.3)$$

where the product is over the ranks R of those observations T'_{0i} , $1 \leq i \leq n$, such that $T'_{0i} \leq t < T_{\max}$ and T'_{0i} corresponds to a death from at least one cause in subset J . If $T_{\max} = T_{ji}$ for some $j \in J$ and some i , $1 \leq i \leq n$, then $\hat{\bar{M}}_J(t)$ is defined to be zero for $t > T_{\max}$. Otherwise, $\hat{\bar{M}}_J(t)$ is undefined for $t > T_{\max}$.

The estimator (4.3) is a generalization of the well-known product-limit estimator for a survival probability proposed by Kaplan and Meier (1958). If $T_{0i} = T_{ji}$ for each i , $1 \leq i \leq n$, and some fixed j , $1 \leq j \leq r$, then (4.3) reduces to a step function with jumps of height $1/n$ at each T_{0i} , thus yielding the usual empirical estimate of $\bar{M}_j(t)$. Assuming (A1), (A2), and (A3), Peterson (1975) shows that the estimator (4.3) is maximum likelihood, (weakly) consistent, and, regarded as a process in t , converges to a normal process.

In the remainder of this section we drop the assumption (A1) of independent risks. How then can we estimate the functions $\bar{M}_I(t)$, $I \in \mathcal{I}$? Note that

formula (4.2) of Theorem 4.2 exhibits a relationship between distributions of observable quantities [viz., the survival probability $\bar{F}(t)$ and the functions $F(t, i)$, $1 \leq i \leq r$] and distributions associated with the theoretical random variables H_i , $i \in I$, which are unobservable. Replacing $\bar{F}(t)$ and $\bar{F}(t, i)$ in (4.2) by their empirical counterparts thus allows us to estimate the distributions G_i , $1 \leq i \leq r$, associated with the unobservable random variables H_i , $1 \leq i \leq r$. Unfortunately, the distributions G_i , $1 \leq i \leq r$, are, in general, different from the marginal distributions M_i , $1 \leq i \leq r$, which we seek to estimate. The natural question then is how to relate the unobservable (but estimable) functions G_i , $1 \leq i \leq r$, to the marginal distributions M_i , $1 \leq i \leq r$. More generally, how can we relate the functions M_I to the survival probabilities \bar{G}_I given by (4.2)?

Assuming no ties (A2) and no common discontinuities among the marginal distributions (A3), Peterson (1975) finds necessary and sufficient conditions for a relationship to exist between the functions \bar{G}_I and \bar{M}_I . Dropping the assumption (A2) of no ties and weakening the assumption (A3) of no common discontinuities, LPQ (1977c) prove the following:

Theorem 4.4. Assume that the functions $F(\cdot, I)$, $i \in I$, in Theorem 4.2 have no common discontinuities. Let $i \in I$. Then for each $t \in [0, \alpha(F))$,

$$\bar{M}_I(t) = \prod_{j \in I, j \neq i} \bar{G}_j(t) \quad (4.4)$$

if and only if the following two conditions hold:

$$\bar{M}_I(a)/\bar{M}_I(a^-) = \begin{cases} \bar{F}(a)/\bar{F}(a^-), & a \in D(F(\cdot, I_i)) \\ 1, & \text{otherwise,} \end{cases} \quad (4.5a)$$

and

$$P(T_{\bar{I}} \geq t | T_i = t) = P(T_{\bar{I}} > t | T_i > t), \quad (4.5b)$$

where $\bar{G}_J(t)$ is given by (4.2) and $D(F(\cdot, I_I))$ is the set of discontinuities of the function $F(t, I_I) = P(T > t, \xi(T) \in J, J \cap I \neq \emptyset)$.

Desu and Narula (1977) arrive at a condition similar to (4.5b) when the assumption of absolute continuity (A4) and hence also the assumption of no ties (A2) hold. We remark that the assumption of no ties (A2) does not hold, e.g., in models where failures or deaths from simultaneous causes can occur. An important family of multivariate distributions for which assumption (A2) fails is the family of multivariate exponential distributions of Marshall and Olkin (1967). We illustrate with an example.

Example 4.5. For simplicity, suppose that the random vector (T_1, T_2) has the Marshall-Olkin bivariate exponential distribution with survival probability:

$$P(T_1 > t_1, T_2 > t_2) = \exp[-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)],$$

for $t_1, t_2 \geq 0$ and $\lambda_1, \lambda_2, \lambda_{12} > 0$. Since the marginal distributions M_1 and M_2 are continuous, condition (4.5a) of Theorem 4.4 holds trivially.

Condition (4.5b) with $I = \{1\}$ states that

$$P(T_2 \geq t | T_1 = t) = P(T_2 > t | T_1 > t).$$

An easy computation shows that $P(T_2 \geq t | T_1 = t) = P(T_2 > t | T_1 > t) = \exp(-\lambda_2 t)$.

Thus, Theorem 4.4 may be applied when the joint distribution belongs to the family of Marshall-Olkin MVE distributions, whereas other approaches to the estimation problem do not apply here since the assumption of no ties (A2) fails to hold.

Formula (4.2), via Theorem 4.4, can now be used to suggest estimators for the functions M_I , $I \in \mathcal{I}$, in the important practical cases when independence fails to hold and when ties are allowed. Suppose that the joint distribution of T_1, \dots, T_r satisfies (4.5 a,b). For each $i = 1, \dots, n$, only the following are observed:

$$T_{0i} = \min(T_{1i}, \dots, T_{ri}),$$

and

$$\xi_i,$$

where $\xi_i = J$ if and only if $T_{0i} = T_{ji}$ for each $j \in J$ and $T_{0i} \neq T_{ji}$ for each $j \notin J$. In accordance with Theorem 4.4, we estimate $\bar{M}_I(t)$ for $t \leq T_{\max}$ by

$$\tilde{\bar{M}}_I(t) = \prod_{J \cap I \neq \emptyset} \tilde{\bar{G}}_J(t),$$

where $\tilde{\bar{G}}_J$ is the function resulting from (4.2) by replacing \bar{F} and $F(\cdot, I)$ by their empirical counterparts, $I \in I$. In analogy with (4.3), the resulting statistic can be expressed as follows:

$$\tilde{\bar{M}}_I(t) = \prod_R [(n - R)/(n - R + 1)], \quad (4.6)$$

where the product is over the ranks R of those observations $T'_{0,R}$, $1 \leq R \leq n$, such that $T'_{0,R} \leq t \leq T_{\max}$ and $T'_{0,R}$ corresponds to a death from the simultaneous causes $j \in J$ such that $J \cap I \neq \emptyset$. If for some i , $T_{\max} = T_{ji}$ for each $j \in J$ with $J \cap I \neq \emptyset$, then $\tilde{\bar{M}}_I(t)$ is defined to be zero for $t > T_{\max}$. Otherwise, $\tilde{\bar{M}}_I(t)$ is undefined for $t > T_{\max}$.

Optimality properties of the estimator (4.6) readily suggest themselves because of its resemblance to the (generalized) Kaplan-Meier estimator (4.3) and its reliance upon empirical distributions. Moreover, there is evidence that formula (4.2) can be used in a similar way to estimate other quantities of interest to the biomedical researcher.

REFERENCES

- [1] Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart, and Winston, New York.
- [2] Birnbaum, Z.W., Esary, J.D., and Saunders, S.C. (1961). Multicomponent systems and structures, and their reliability. Technometrics 3, 55-77.
- [3] Desu, M.M. and Narula, S.C. (1977). Reliability estimation under competing causes of failure. The Theory and Applications of Reliability with Emphasis on Bayesian and Nonparametric Methods, I. Shimi and C.P. Tsokos, ed., Academic Press. In press.
- [4] Esary, J.D. and Marshall, A.W. (1970). Coherent life functions. SIAM J. Appl. Math. 18, 810-814.
- [5] Esary, J.D. and Marshall, A.W. (1974). Multivariate distributions with exponential minimums. Ann. Statist. 2, 84-98.
- [6] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.
- [7] Kotz, S. (1974). Annotated and abstracted bibliography on multivariate statistical models. Temple University Technical Report.
- [8] Langberg, N., Proschan, F., and Quinzi, A.J. (1977a). Converting dependent models into independent ones, with applications in reliability. The Theory and Applications of Reliability, with Emphasis on Bayesian and Nonparametric Methods, I. Shimi and C.P. Tsokos, ed., Academic Press. In press.
- [9] Langberg, N., Proschan, F., and Quinzi, A.J. (1977b). Converting dependent models into independent ones, preserving essential features. Submitted for publication.
- [10] Langberg, N., Proschan, F., and Quinzi, A.J. (1977c). Dependent competing risks. In process.
- [11] Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. J. Amer. Statist. Assoc. 62, 30-44.
- [12] Miller, D.R. (1977). A note on independence of multivariate lifetimes in competing risks models. Ann. Statist., in press.
- [13] Peterson, A.V., Jr. (1975). Nonparametric estimation in the competing risks problem. Stanford University Technical Report No. 13.
- [14] Proschan, F. and Sullo, P. (1976). Estimating the parameters of a multivariate exponential distribution, J. Amer. Statist. Assoc. 71, 465-472.
- [15] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. USA 72.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M399 AFOSR No. 69	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE Transformations Yielding Reliability Models Based on Independent Random Variables: A Survey	5. TYPE OF REPORT & PERIOD COVERED Technical Report	6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M399
7. AUTHORS(s) N. Langberg F. Proschan A. J. Quinzi	8. CONTRACT OR GRANT NUMBER(s) AFOSR-74-2581C	
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME & ADDRESS The U. S. Air Force Air Force Office of Scientific Research Bolling Air Force Base, D. C. 20332	12. REPORT DATE January, 1977	13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report) Unclassified	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS Reliability Independent random variables Dependent random variables Exponential minima Shock models Lifelength Survey paper Multivariate exponential distribution Competing risks		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

20. ABSTRACT

This is an expository paper presenting various ways of transforming dependent models into independent ones and displaying applications in a variety of contexts including reliability modelling, life testing, and nonparametric estimation in the study of competing risks.

Transformations Yielding Reliability Models Based
on Independent Random Variables: A Survey

by

N. Langberg¹, F. Proschan¹, and A.J. Quinzi²

January, 1977

¹Florida State University

²Florida State University and Temple University

Research supported by the Air Force Office of Scientific Research under
AFOSR Grant 74-2581C.