The Annals of Statistics
1977, Vol. 5, No. 3, 576-579

A NOTE ON INDEPENDENCE OF MULTIVARIATE LIFETIMES IN COMPETING RISKS MODELS

BY DOUGLAS R. MILLER

University of Missouri, Columbia

Given the joint distribution of cause of failure and system lifetime of a series system in which components fail simultaneously with probability 0, it is possible (under mild regularity conditions) to assume a model for the system in which component lifetimes are independent.

1. Introduction and summary. A system is subject to competing risks if failure of the system may result from each of k causes—for example, a biological organism susceptible to k different fatal diseases or a series system consisting of k components. It is assumed that system failure is due to exactly one of the causes. It is possible to observe the system lifetime, L, and the failure pattern, I, the index of the cause of failure. The joint distribution $F_{L,I}$ is assumed to satisfy the following requirement:

HYPOTHESIS A. The conditional distributions $F_{L|I=i}$, $i = 1, 2, \cdots, k$, have no common atoms.

The purpose of this paper is to show that under Hypothesis A it is possible to use *independent* component lifetimes as a model for this system, namely:

THEOREM 1. *Let $F_{L,I}$ be a probability distribution on* $[0, \infty) \times \{1, 2, \cdots, k\}$ *which satisfies Hypothesis* A. *Then there exist independent random variables T_1, $T_2, \cdots, T_k$, such that* $(\min(T_1, T_2, \cdots, T_k), i(T_1, T_2, \cdots, T_k))$ *has distribution $F_{L,I}$, where $i(t_1, t_2, \cdots, t_k) = j$ if $t_j = \min(t_1, t_2, \cdots, t_k)$. Furthermore, at least one of $T_1, T_2, \cdots, T_k$ has a proper distribution; and the distributions are uniquely defined on* $[0, \inf\{t : F_L(t) = 1\})$. (Note that $i(\cdot)$ need not be defined in the case of "ties.")

Before proving Theorem 1 in Section 2, several remarks are in order:

(i) This research was performed without the knowledge that Tsiatis (1975) had recently proved Theorem 1 under the assumption of absolutely continuous system lifetime distributions. Subsequently Peterson (1975) studied the problem, arriving independently at Theorem 1. Most recently Langberg, Proschan and Quinzi (1976) have done a rather exhaustive analysis, considerably extending Theorem 1. Their proofs are all based on failure-rate functions. The proof in this paper differs considerably: it is based on an attempt to find a test for independence.

(ii) There may be other considerations (such as restrictions to certain

Received April 1975; revised September 1976.

AMS 1970 *subject classifications.* Primary 62N05; Secondary 62H05.

Key words and phrases. Reliability theory, series system, multivariate lifetime distribution, competing risks model.

parametric families, see Moeschberger and David (1971)) which make models assuming dependent lifetimes attractive or necessary even when by Theorem 1 it is not mathematically necessary; see also the discussion in Barlow and Proschan (1975), Chapter 5.

(iii) If simultaneous failures of components can occur with positive probability, then the conclusion of the theorem need not hold and a model with dependent lifetimes is necessary. In this instance, it is interesting that Marshall and Olkin's (1966) bivariate exponential distribution consists of a component corresponding to the joint distribution of independent exponentials plus a singular component corresponding to simultaneous failures. They show this distribution to be a much more natural model in reliability theory than are other bivariate exponential distributions.

(iv) A word should be said about the possibility of improper random variables in Theorem 1: using Tsiatis' (1975) approach, it can be shown that, for $k = 2$, if $P(L > t) = \exp(-t)$ and $P(I = 1 \mid L = t) = (1 + t)^{-2}$, then $P(T_1 \leqq t_1) = 1 - \exp(-\int_0^{t_1} (1 + s)^{-2}\, ds)$; thus $P(T_1 < \infty) = 1 - e^{-1}$. It would be interesting to characterize the distributions $F_{L,I}$ for which $T_1, \cdots, T_k$ all have proper distributions. One conjecture is that if L and I are independent, then $T_1, T_2, \cdots, T_k$ are proper.

2. Proof of Theorem 1. The proof of Theorem 1 for the case $k = 2$ illustrates the general idea without cumbersome notation required for arbitrary k. The proof consists of 3 lemmas. Lemma 2 is the crux. Lemmas 1 and 3 involve only approximating a general distribution $F_{L,I}$ by a discrete distribution.

LEMMA 1. *Let $F_{L,I}$ be a probability distribution on $[0, \infty) \times \{1, 2\}$ such that $F_{L|I=1}$ and $F_{L|I=2}$ have no common atoms. Then, given $\varepsilon > 0$, there exists a set*

$$\begin{aligned} &\Lambda = \{\{s_1, s_2, \cdots, s_n\} \times \{1\}\} \cup \{\{t_1, t_2, \cdots, t_n\} \times \{2\}\}\,, \\ (2.1)\qquad &0 \leqq s_1 < s_2 < \cdots < s_n\,, \quad 0 \leqq t_1 < t_2 < \cdots < t_n\,, \\ &s_i \neq t_j\,, \quad 1 \leqq i, j \leqq n\,. \end{aligned}$$

and a probability distribution $G_{L,I}$ with support Λ such that $\sup |F_{L,I} - G_{L,I}| \leqq \varepsilon$.

PROOF. The proof is a routine exercise in approximating a distribution function by a discrete one.

LEMMA 2. *Let $G_{L,I}$ be a distribution supported on Λ, (2.1). Then there exist unique independent random variables S and T taking values $[s_i, i = 1, \cdots, n\}$ and $\{t_i, i = 1, \cdots, n\}$, respectively, (at least one of which has a proper distribution) such that* $(\min(S, T), i(S, T))$ *has distribution $G_{L,I}$.*

PROOF. Let $G_{L,I}(s_i, 1) = a_i$ and $G_{L,I}(t_i, 2) = b_i$, $i = 1, 2, \cdots, n$. Combine $\{s_i, i = 1, \cdots, n\}$ and $\{t_i, i = 1, \cdots, n\}$ into one ordered set $\{u_i, i = 1, \cdots, 2n\}$; in particular $u_i = s_j$ if $t_{i-j} < s_j < t_{i-j+1}$. Define the distribution of independent random variables S and T, $P\{S = s_i\} = p_i$, $P\{T = t_i\} = q_i$, $i = 1, \cdots, n$, as follows: If $u_1 = s_1$, let $p_1 = a_1$; then, if $u_2 = t_1$, in order that $P\{S \geqq s_2, T = t_1\} = P\{S \geqq s_2\}P\{T = t_1\}$ it is necessary that $b_1 = (1 - p_1)q_1$, so define $q_1 = b_1/(1 - p_1)$.

If $u_1 = t_1$, let $q_1 = b_1$. Continuing formally by induction: If $u_i = s_j$, then p_k, $k = 1, \cdots, j-1$, and q_k, $k = 1, \cdots, i-j$ have been defined. In order for independence of S and T, $a_j = P\{S = s_j, T \geqq t_{i-j+1}\} = P\{S = s_j\} \cdot P\{T \geqq t_{i-j+1}\} = p_j(1 - \sum_{k=1}^{i-j} q_k)$. Thus define $p_j = a_j/(1 - \sum_{k=1}^{i-j} q_k)$. If $u_i = t_j$, then define q_j similarly. By the above arguments it follows by induction that for any sequences, $s_1 < s_2 < \cdots < s_n$ and $t_1 < t_2 < \cdots < t_n$, when pooled to form $u_1 < u_2 < \cdots < u_{2n}$, that

$$1 - \sum_{k=1}^{i} a_k - \sum_{k=1}^{j-i} b_k = (1 - \sum_{k=1}^{i} p_k)(1 - \sum_{k=1}^{j-i} q_k)\,, \quad j = 1, \cdots, 2n\,, \tag{2.2}$$

if $\{u_1, \cdots, u_j\} = \{s_1, \cdots, s_i\} \cup \{t_1, \cdots, t_{j-i}\}$. The left-hand side of (2.2) is nonnegative and monotonically decreasing as $j \uparrow 2n$ and consequently so is the right-hand side. This proves that $p_i \geqq 0$, $\sum_{k=1}^{i} p_k \leqq 1$, $q_i \geqq 0$, and $\sum_{k=1}^{i} q_k \leqq 1$, for $i = 1, 2, \cdots, n$. Thus $\{p_i, i = 1, \cdots, n\}$ and $\{q_i, i = 1, \cdots, n\}$ are indeed (possibly improper) probabilities. The left-hand side of (2.2) equals 0 when $j = 2n$; thus at least one of S or T must be proper. It follows by the above construction that, if S and T are independent, $(\min(S, T), i(S, T))$ has the distribution $G_{L,I}$, completing the proof.

For any $0 < s_1 < s_2 < \cdots < s_n$ and $0 < t_1 < t_2 < \cdots < t_n$, there exist $2n$ sets of the form $\{S = s_i, T > t_j\}$, where $t_{j-1} < s_i < t_j$, or $\{T = t_k, S > s_m\}$, where $s_{m-1} < t_k < s_m$. These sets partition the plane into $2n$ cells, which would correspond to "observable" events in a contingency table for testing the independence of S and T; cf. diagram 2.1 of Rose (1973). The $2n$ equations of (2.2) each constitute a constraint imposed by an independent model. This leaves $2n - 2n = 0$ degrees of freedom to test the hypothesis of independence; roughly speaking, this is equivalent to the fact that an independent model fits the data perfectly.

LEMMA 3. *Let the random variables L and I have joint distribution $F_{L,I}$ which satisfies Hypothesis* A. *Assume, for $\varepsilon > 0$, there exist independent random variables S_ε and T_ε such that* $\sup |F_{\min(S_\varepsilon, T_\varepsilon), i(S_\varepsilon, T_\varepsilon)} - F_{L,I}| < \varepsilon$. *Then there exist independent random variables S and T such that $(S_\varepsilon, T_\varepsilon) \to_D (S, T)$ as $\varepsilon \to 0$ and* $(\min(S, T), i(S, T)) =_D (L, I)$. *The random variables S and T are uniquely determined on* $[0, \inf\{t : F_L(t) = 1\})$ *and at least one of them has a proper distribution.*

PROOF. The Helly selection theorem ([2], page 227) implies the existence of a convergent subsequence with limit (S, T). Independence follows from independence of $(S_\varepsilon, T_\varepsilon)$. The continuous mapping theorem ([2], page 30) implies that $(\min(S, T), i(S, T)) =_D (L, I)$. The proper distribution of L and the independence of S and T together imply that either S or T must have a proper distribution. Uniqueness is proved by contradiction.

REFERENCES

[1] BARLOW, R. E. and PROSCHAN, F. (1975). *The Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston, New York.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] LANGBERG, N., PROSCHAN, F. and QUINZI, A. J. (1976). Converting dependent models into independent ones, preserving essential features. Florida State Univ. Statistics Report M367.

[4] MARSHALL, A. W. and OLKIN, I. (1966). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62** 30-44.

[5] 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.

[6] PETERSON, A. V. (1975). Nonparametric estimation in the competing risks problem. Technical report no. 73 (September 10, 1975), Department of Statistics, Stanford Univ.

[7] ROSE, D. M. (1973). An investigation of dependent competing risks. Ph. D. dissertation, Univ. of Washington.

[8] TSIATIS, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. USA* **72** 20-22.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65201