FAULT DIVERSITY IN SOFTWARE RELIABILITY

by

Philip J. Boland¹, Frank Proschan² and Y. L. Tong³

FSU Statistics Report No. M714
AFOSR Technical Report No. 85-185

September 1985

University College, Dublin
Department of Mathematics
Belfield, Dublin 4, Ireland

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

and

Georgia Institute of Technology
School of Mathematics
Atlanta, Georgia 30332

AMS Subject Classification Numbers: 68M15, 62N05, 60E15,
26A51.

Key Words: Software reliability, faults or bugs in a system,
Jelinski-Moranda model, fault diversity, majorization, multinomial distribution.

¹,² Research supported by AFOSR no. 82-K-007.

³ Research supported in part by NSF no. DMS-8502346 and
AFOSR no. 82-K-007.
Fault Diversity in Software Reliability

by

P. J. Boland, F. Proschan and Y. L. Tong

Abstract

Diversity of bugs or faults in a software system is a factor contributing to software unreliability which has not yet been appropriately emphasized. This paper is written with the intention of demonstrating the impact of fault diversity on the time to detection of software bugs. A new discrete software reliability model based on the multinomial distribution is introduced. It is shown that for models of this type, the more diverse the fault probabilities are, the longer (stochastically) it takes to detect or eliminate any n faults, while the smaller (stochastically) will be the number of faults detected or eliminated during a given amount of time (or during a given number of inputs to the system). The impact of fault diversity is also demonstrated for the Jelinski-Moranda model.
§1. Introduction and Summary

The abundance and wide use of computer software has led to considerable research in software reliability. One could define software reliability as the probability that a computer programme or software package functions without failure under specified conditions during an assigned time or mission period. For a given software package, one might also be interested in estimating the fault content or total number of bugs in the 'system.' The distribution of the time to detection of any n bugs or the distribution of the number of faults detected during a specified period of operation would also be valuable in assessing the performance capabilities of a software package. Many probabilistic models have been developed in an effort to estimate these and other measures of software reliability.

Perhaps the most widely referenced model in software reliability is that proposed by Jelinski and Moranda (1972). They assume that initially there are an unknown number of faults $N$ in the software system and that at any point in time the failure rate of the system is proportional to the number of faults still resident in the system. They furthermore assume that if $\tau_i$ is the random variable denoting the time between discovery of $i-1$ and $i$ faults or bugs, then the $\tau_i$ are independent exponential random variables with parameters $\lambda_i = (N-i+1)\lambda$ for some proportionality constant $\lambda > 0$, $i = 1, \ldots, N$. Note that in the Jelinski-Moranda model
each fault in the system has the same effect on the overall failure rate of the system. In fact another interpretation of this model is that there are an unknown number $N$ of faults in the system, the 'appearance' times (detection times) of which are independently and exponentially distributed with (unknown) parameter $\lambda$.

The simplicity of the Jelinski-Moranda model makes it a particularly appealing one, although it is not without its critics. Using the Jelinski-Moranda model, Forman and Singpurwalla (1977) propose a stopping rule for debugging computer software. They note that the maximum likelihood estimate $\hat{N}$ of the total fault content $N$, based on a sample of $n < N$ fault appearance times, is unstable and often produces misleading results. Goudie and Goldie (1981) and Joe and Reid (1985) show respectively that in fact the maximum likelihood estimates $\hat{N}$ of $N$ based on either a sample of the first $n$ fault appearance times or a sample of fault appearance times during a fixed interval of time $[0,t_0]$ can be infinite with substantial probability.

Littlewood and Verrall (1973) were also critical of the Jelinski-Moranda model in arguing that to assume each fault or bug makes the same contribution to the overall failure rate is not realistic, and that furthermore there is some uncertainty in the removal of a fault. They hence modify the Jelinski-Moranda model in assuming that the inter-arrival times between faults $\tau_i$ are independent exponential
random variables with parameters \( \lambda_i \), \( i = 1, \ldots, N \), but where the \( \lambda_i \)'s are themselves random variables and the sequence \( \{ \lambda_i \} \) is stochastically decreasing in \( i \).

Goel and Okumoto (1979) consider a software reliability model of the Poisson type. They let \( M(t) \) denote the cumulative number of software failures in \([0, t]\), and under appropriate assumptions \( \{ M(t) : t \geq 0 \} \) is a nonhomogeneous Poisson process where \( E(M(t)) = a(1 - e^{-bt}) \), \( a \) and \( b \) being unknown parameters. Reliability estimates based on this model tend to be more conservative than those of the Jelinski-Moranda model. Musa and Okumoto (1984) introduce a logarithmic Poisson model. They assume that \( \{ M(t) : t \geq 0 \} \) is a nonhomogeneous Poisson process with failure intensity function \( M'(t) \) of the form \( M'(t) = m_0 e^{-\theta M(t)} \), where \( m_0 \) and \( \theta \) are unknown parameters. When the maximum likelihood estimates of \( m_0 \) and \( \theta \) exist, Musa and Okumoto claim this model possesses good predictive properties.

Many other software reliability models have been proposed, all of course with their own 'faults.' Langberg and Singpurwalla (1985) have attempted to unify some of these models into a general framework.

Research in software reliability attempts to determine those factors which contribute to software unreliability and to provide means for assessing these factors. One factor that we feel has not received enough attention in the literature is that of bug or fault diversity. This
paper is written with the intention of demonstrating the impact of fault diversity on the time to detection of software bugs. We introduce in Section 2 a new discrete software reliability model based on the multinomial distribution. In Section 3 we review the concept of majorization and prove some basic inequalities for the multinomial distribution. These results are then used in Section 4 to show the impact of fault diversity in the newly developed discrete software reliability model. For instance it is shown in Theorem 4.3 that the more diverse the fault probabilities are in a system, the longer (stochastically) it takes to detect or eliminate n of them, while the smaller (stochastically) will be the number of faults detected or eliminated during a given amount of time (or during a given number of inputs to the system). In Section 5 we similarly demonstrate the impact of fault diversity in the Jelinski-Moranda model.

§2. A New Discrete Software Reliability Model

We now propose a new discrete model for the debugging process of a software system with a finite number N of faults or bugs. We label the model as discrete since we consider as our basic unit of 'time' any input (or demand) made upon the software system or package. In continuous time models like the Jelinski-Moranda model, 'time' usually refers to execution (CPU) or calendar time. In our discrete model, we will assume that each input to the software
system results in either an error free response or in the
detection of precisely one of the bugs or faults still in
the system. We will also assume that inputs to the system
occur independently of one another.

If the system originally contains $N$ bugs, we denote
bug $i$ by $b_i$ for $i = 1, \ldots, N$. The initial 'state' of the
software system can be described by a vector $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$
where $\theta_i$ is the probability that bug $b_i$ is encountered or
detected as the result of an input to the system, $i = 1, \ldots, N$, and $\theta_0$ is the initial probability that an input
to the system will result in no bug detection. We assume
that $\theta_i > 0$ for $i = 1, \ldots, N$ and $\sum_{i=0}^{N} \theta_i = 1$. In most prac-
tical situations $\theta_0$ will be considerably larger than $\sum_{i=1}^{N} \theta_i$.

Let us now assume that a positive integer $k$ exists
such that any bug $b_i$ is permanently eliminated from the
system immediately after it is encountered or detected for
the $k^{th}$ time. In most situations $k = 1$. We assume that
no new bugs are introduced into the system at any stage.
Of course the probabilities of bug detection and fault free
input response change each time a bug is eliminated from the
system. Given the initial vector $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$, let
us denote by $\theta_{i}^{j}$ the probability that bug $b_i$ is encountered
on the $j^{th}$ input to the system. Then the following induct-
tive relations hold:

$$\theta_{i}^{1} = \theta_i \quad \text{for } i = 0, 1, \ldots, N,$$
\[ \theta_{i}^{j+1} = \begin{cases} 0 & \text{if bug } b_{i} \text{ is encountered for the } k^{th} \text{ time on input } j \\ \theta_{i}^{j} & \text{otherwise} \quad \text{for } i = 1, \ldots, N \text{ and } j \geq 1, \end{cases} \]

and

\[ \theta_{0}^{j+1} = \begin{cases} \theta_{0}^{j} + \theta_{i} & \text{if bug } b_{i} \text{ is encountered for the } k^{th} \text{ time on the } j^{th} \text{ input for some } i \\ \theta_{0}^{j} & \text{otherwise} \quad \text{for } j \geq 1. \end{cases} \]

With \( \theta = (\theta_{0}, \theta_{1}, \ldots, \theta_{N}) \) and \( k \) given, we are particularly interested in the random variables:

(1) \( T_{n} = \) time (number of inputs into the system) until any \( n \) of the \( N \) bugs are eliminated.

and (2) \( B^{m} = \) number of bugs eliminated from the system after \( m \) inputs into the system.

We will demonstrate the impact on the random variables \( T_{n} \) and \( B^{m} \) made by the diversity among the various bug probabilities \( \theta_{1}, \ldots, \theta_{N} \). We will see that the more diverse \((\theta_{1}, \ldots, \theta_{N})\) is, the larger \( T_{n} \) is stochastically (that is the longer it will take stochastically to eliminate \( n \) bugs from the system) while the smaller \( B^{m} \) is stochastically (that is the fewer the number of bugs eliminated will be stochastically after \( m \) inputs).
§3. Majorization Applied to the Multinomial Distribution

Definition 3.1. For \( u = (u_1, \ldots, u_N) \in \mathbb{R}^N \), we let

\( u^+ = (u_{[1]}, \ldots, u_{[N]}) \) be the vector with the components of \( u \) arranged in decreasing order. We say that the vector \( u \) majorizes \( u' \) (written \( u >^m u' \)) if

\[
\sum_{i=1}^{n} u_{[i]} \geq \sum_{i=1}^{n} u'_{[i]} \quad \text{for } n = 1, \ldots, N-1
\]

and

\[
\sum_{i=1}^{N} u_{[i]} = \sum_{i=1}^{N} u'_{[i]}.
\]

When \( u \) majorizes \( u' \), the coordinates of \( u \) are more 'dispersed' than those of \( u' \). A function \( f \) which reflects the preordering \( >^m \) of majorization on \( \mathbb{R}^N \) in that \( u >^m u' \Rightarrow f(u) \geq f(u') \) is called a Schur convex function. See Marshall and Olkin (1979) for an excellent account of majorization and Schur functions.

Definition 3.2. (Nevius, Proschan and Sethuraman (1977)).

Let \( Y \) and \( Y' \) be random vectors taking values in \( \mathbb{R}^N \). We say that \( Y \) stochastically majorizes \( Y' \) (written \( Y \geq^\text{st.m.} Y' \)) if \( f(Y) \geq^\text{st} f(Y') \) (that is \( P[f(Y) \geq c] \geq P[f(Y') \geq c] \) for all \( c \)) for all Borel measurable Schur convex functions \( f \) on \( \mathbb{R}^N \).

For \( \theta = (\theta_1, \ldots, \theta_N) \in \mathbb{R}^N \) (respectively \( \theta = (\theta_0', \theta_1', \ldots, \theta_N') \in \mathbb{R}^{N+1} \)) and \( m \) a positive integer, we let \( Y^m_\theta = (Y_1, \ldots, Y_N) \) (respectively \( X^m_\theta = (X_0, X_1, \ldots, X_N) \)) be the multinomial random vector with parameters \( m \) and \( \theta \). Nevius,
Proschan and Sethuraman (1977) have shown that the multinomial family of distributions is a Schur family in the sense that $\bar{\theta} \overset{m}{\succ} \bar{\theta}' \Rightarrow Y^m_{\bar{\theta}} \overset{\text{st.m.}}{\geq} Y^m_{\bar{\theta}'}$. This fact will be an important tool in our consideration of multinomial distributions, as evidenced by the following lemma:

**Lemma 3.3.** For each $k \geq 0$, $m \geq 1$ and $\bar{\theta} = (\theta_1, \ldots, \theta_N)$ where $1 = \sum_{i=1}^{N} \theta_i$ and $\theta_i \geq 0$ for $i = 1, \ldots, N$, we denote by

$$M^m_{\bar{\theta}}, k = P(Y_1 \geq k, \ldots, Y_N \geq k; Y = (Y_1, \ldots, Y_N) = Y^m_{\bar{\theta}})$$

$$= P_{\bar{\theta}}(Y_1 \geq k, \ldots, Y_N \geq k).$$

Then

a) $\bar{\theta} \overset{m}{\succ} \bar{\theta}' \Rightarrow M^m_{\bar{\theta}}, k \leq M^m_{\bar{\theta}'}$, $k$

and

b) $M^m_{\bar{\theta}}, k$ is an increasing function of $m$ for fixed $k$ and $\bar{\theta}$.  

**Proof.** a) This is a consequence of the fact that the multinomial family is a Schur family together with the fact that the indicator function of the set \{$(y_1, \ldots, y_N)$: $y_i \geq k$, $i = 1, \ldots, N$\} in $\mathbb{R}^N$ is Schur convex (see Application 4.4 of Nevius, Proschan and Sethuraman (1977)).

b) It is clear that for a given $\bar{\theta}$, each of the events $Y_i \geq k$ for $i = 1, \ldots, N$ happening simultaneously is more likely in $m+1$ trials than in $m$, and hence the result.
The following lemma will be useful in the proof of Theorem 3.5:

**Lemma 3.4.** If \( \{f(x; \theta) : \theta \in \Omega \subset \mathbb{R} \} \) is a family of densities or frequency functions with the monotone likelihood ratio property, then \( E_{\theta} \phi(X) \) is nondecreasing in \( \theta \) for all non-decreasing functions \( \phi \), provided the expectations exist.

**Proof.** See for example Chapter 3 of Lehmann (1959).

**Theorem 3.5.** Let \( X = (X_0, X_1, \ldots, X_N) \) be a multinomial random vector with values in \( \mathbb{R}^{N+1} \). Let \( \theta = (\theta_0, \theta_1, \ldots, \theta_N) \) and \( \theta' = (\theta_0', \theta_1', \ldots, \theta_N') \) be such that

a) \( \theta_0 \geq \theta_0' \)

and b) \( \left( \frac{\theta_1}{1-\theta_0}, \ldots, \frac{\theta_N}{1-\theta_0} \right) \geq^m \left( \frac{\theta_1'}{1-\theta_0'}, \ldots, \frac{\theta_N'}{1-\theta_0'} \right) \).

Then for any \( m \) and \( k \)

\[
P_{\theta}(X_1 \geq k, \ldots, X_N \geq k) \leq P_{\theta'}(X_1 \geq k, \ldots, X_N \geq k).
\]

**Proof.**

\[
P_{\theta}(X_1 \geq k, \ldots, X_N \geq k) = \sum_{j=0}^{m} P_{\theta}(X_1 \geq k, \ldots, X_N \geq k | X_0 = j) P_{\theta}(X_0 = j)
\]

\[
= \sum_{j=0}^{m} M^{m-j} \left( \frac{\theta_1}{1-\theta_0}, \ldots, \frac{\theta_N}{1-\theta_0} \right)^{j} \left( \frac{m-j}{1-\theta_0'} \right)^{j} \left( \frac{m-j}{1-\theta_0} \right)^{m-j}
\]

\[
\leq \sum_{j=0}^{m} M^{m-j} \left( \frac{\theta_1'}{1-\theta_0'}, \ldots, \frac{\theta_N'}{1-\theta_0'} \right)^{j} \left( \frac{m-j}{1-\theta_0'} \right)^{j} \left( \frac{m-j}{1-\theta_0} \right)^{m-j}
\]

(by Lemma 3.3a)
\[ \leq \sum_{j=0}^{m} \frac{M^{m-j} \theta_0^j \theta_1^0 \cdots \theta_N^j (1-\theta_0)^j (1-\theta_1)^{m-j} \cdots (1-\theta_N)^{m-j}}{1-\theta_0^j}, k \quad \text{(by Lemma 3.4)} \]

\[ = P_{\theta'}(X_1 \geq k, \ldots, X_N \geq k). \]

More generally the proof of Theorem 3.5 may be readily adapted to yield the following corollary:

**Corollary 3.6.** Let \( X = (X_{01}, \ldots, X_{0l}, X_1, \ldots, X_N) \) be a random vector with values in \( \mathbb{R}^{l+N} \). Let \( \theta = (\theta_{01}, \ldots, \theta_{0l}, \theta_1, \ldots, \theta_N) \)
and \( \theta' = (\theta'_0, \ldots, \theta'_0, \theta'_1, \ldots, \theta'_N) \) be such that

a) \( \theta_0 = \theta_{01} + \cdots + \theta_{0l} \geq \theta'_0 + \cdots + \theta'_0 = \theta'_0 \)

and b) \( \left( \frac{\theta_1}{1-\theta_0}, \ldots, \frac{\theta_N}{1-\theta_0} \right) \leq m \left( \frac{\theta'_1}{1-\theta'_0}, \ldots, \frac{\theta'_N}{1-\theta'_0} \right) \).

Then for any \( m \) and \( k \), \( P_{\theta}(X_1 \geq k, \ldots, X_N \geq k) \leq P_{\theta'}(X_1 \geq k, \ldots, X_N \geq k) \).

§4. **Discrete Software Reliability Modelling Using the Multinomial Distribution**

We return to our discrete software reliability model, where initially (and until some bug or fault is removed from the system) the probability of encountering bug \( b_i \) on an input to the system is \( \theta_i \) for \( i = 1, \ldots, N \) while with probability \( \theta_0 \) no bug is encountered. Let us now denote by \( A_{k,n,\theta}^m \) (we will write \( A_{k,n}^m \) when the value of \( \theta \) is clear) the event that in \( m \) independent inputs to the system, \( n \) or more bugs are encountered \( k \) times each (that is, \( n \) or more bugs are eliminated from the system since \( k \) is the threshold value such that when a bug is observed for the \( k \)th time it
is corrected and one less fault exists in the system).

The probability of the event $A_{k,n}^m$ for a given $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$ can be very tedious to calculate directly. For example, an expression for the probability of $A_{1,n}^m$ is:

$$P(A_{1,n}^m) = \min(m,N) \sum_{r=n} \sum_{\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, N\}} \theta_{i_1} \cdots \theta_{i_r} \sum_{\pi \text{ permutation of } \{1, \ldots, r\}} \theta_0^{s_0} (\theta_0 + \theta_{i_{\pi(1)}})^{s_1} \cdots (\theta_0 + \theta_{i_{\pi(r)}})^{s_r}$$

$$s = (s_0, s_1, \ldots, s_r), \ s_0 + s_1 + \ldots + s_r = m - r$$

while

$$P(A_{1,n}^m) = P(\text{all } N \text{ bugs are encountered once and then eliminated in } m \text{ inputs to the system})$$

$$= \theta_1 \cdots \theta_N \sum_{\pi} \sum_{s=(s_0, s_1, \ldots, s_N)}^{s_0 + \ldots + s_N = m-N} \theta_0^{s_0} (\theta_0 + \theta_{i_{\pi(1)}})^{s_1} \cdots (\theta_0 + \theta_{i_{\pi(N)}})^{s_N}$$

Now for a given $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$ and $m$, let us consider again the multinomial random vector $X^m_{\theta} = (X_0, X_1, \ldots, X_N) = X$ and define $E_{k,n}^m\theta$ (we write $E_{k,n}^m$ when the value of $\theta$ is clear) to be the event that in $m$ trials $n$ or more of $X_1, \ldots, X_N$ takes values $\geq k$. A key observation is that for a given $\theta$, the debugging event $A_{k,n}^m$ corresponds in a natural way to the multinomial event $E_{k,n}^m\theta$, and that in fact they have equal
probabilities.

Lemma 4.1. \( P(A^m_{k,n}) = P(E^m_{k,n}) \).

**Proof.** For any subset \( \{i_1, \ldots, i_r\} \) of \( \{1, \ldots, N\} \) of size \( r \geq n \), we let \( A^m_{k,n} \{i_1, \ldots, i_r\} \) be the event that in \( m \) inputs to the software package bugs \( b_{i_1}, \ldots, b_{i_r} \) and only these bugs are each observed \( k \) times (and hence eliminated). We let \( E^m_{k,n} \{i_1, \ldots, i_r\} \) denote the event \( [X_i \geq k \text{ for all } i \in \{i_1, \ldots, i_r\}, \ X_i < k \text{ for all } i \not\in \{0, i_1, \ldots, i_r\} | X = X_0^m] \). Note that \( E^m_{k,n} \{i_1, \ldots, i_r\} \) is the event for the multinomial random vector \( X_0^m \) where (ignoring \( X_0 \)) the observations of types \( i_1, \ldots, i_r \) (and none others) are made at least \( k \) times.

It is clear that \( A^m_{k,n} \{i_1, \ldots, i_r\} \) and \( E^m_{k,n} \{i_1, \ldots, i_r\} \) correspond in a natural way and have equal probability. The lemma follows from this observation upon taking the disjoint union over all subsets \( \{i_1, \ldots, i_r\} \subset \{1, \ldots, N\} \) where \( r \geq n \).

Definition 4.2. Let \( \theta = (\theta_0, \theta_1, \ldots, \theta_N) \) be an 'initial distribution' for the \( N \) bugs or faults in our software system. Furthermore, suppose that a bug is eliminated after it is encountered for the \( k^{th} \) time, and that when such a bug is eliminated the probability for a no-bug encounter is appropriately modified. For our model, the following random variables are of key interest:

1. \( T_{n, \theta} \) = number of inputs (time) until any \( n \) bugs in the system are eliminated.
and (2) $B^m_\theta$ = number of bugs eliminated after m inputs to the software system.

It is clear that $T_{n, \theta} \geq m+1 \iff B^m_\theta < n$, or in other words that n bugs are eliminated from the system only after m+1 inputs if and only if the number of bugs eliminated from the system after m inputs is less than n.

**Theorem 4.3.** Let $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$ and $\theta' = (\theta'_0, \theta'_1, \ldots, \theta'_N)$ be such that

\[ i) \quad \theta_0 \geq \theta'_0 \]

and \[ ii) \quad \left( \frac{\theta_1}{1-\theta_0}, \ldots, \frac{\theta_N}{1-\theta_0} \right) \succ^m \left( \frac{\theta'_1}{1-\theta'_0}, \ldots, \frac{\theta'_N}{1-\theta'_0} \right). \]

Then a) $T_{n, \theta} \succeq st T_{n, \theta'}$ for any n

and b) $B^m_\theta \leq st B^m_\theta'$ for any m.

**Proof.** Due to the relationship between $T_{n, \theta}$ and $B^m_\theta$, it suffices to prove a).

Now for any m,

\[ P(T_{n, \theta} \geq m+1) = P(B^m_\theta < n) \]

\[ = 1 - P(A^m_k, n, \theta) \]

\[ = 1 - P(E^m_k, n, \theta) \] (by Lemma 4.1)

\[ \geq 1 - P(E^m_k, n, \theta') \] (by Theorem 3.5)

\[ = P(T_{n, \theta'} \geq m+1), \]

and hence $T_{n, \theta} \succeq st T_{n, \theta'}$. 

Remark 4.4. Suppose we compare two software systems where for system 1, \( \vec{\theta} = (0.90, 0.04, 0.04, 0.02) \) while for system 2, \( \vec{\theta}' = (0.80, 0.08, 0.06, 0.06) \). We might say that initially system 1 is 90% fault free, while system 2 is 80% fault free. Now \( \vec{\theta} \) and \( \vec{\theta}' \) satisfy the conditions of Theorem 4.3, and hence in particular system 1 will stochastically take longer than system 2 to debug assuming that the same rule (that is the same value of \( k \)) for eliminating bugs holds for each system. This example illustrates the basic concept that for two systems determined by \( \vec{\theta} = (\theta_0, \theta_1, \ldots, \theta_N) \) (system 1) and \( \vec{\theta}' = (\theta_0', \theta_1', \ldots, \theta_N') \) (system 2), if \( \theta_0 \geq \theta_0' \) and the initial bug probabilities in system 1 are relatively more 'dispersed' than those in system 2, then system 1 will take longer stochastically to eliminate any \( n \) bugs than will system 2; equivalently, the probability that \( n \) or more bugs have been eliminated from system 1 after \( m \) inputs is less than or equal to the probability that \( n \) or more bugs have been eliminated from system 2 after \( m \) inputs -- for any \( m \).

Remark 4.5. Note that if we are comparing two vectors

\[ \vec{\theta} = (\theta_0, \theta_1, \ldots, \theta_N) \] and \( \vec{\theta}' = (\theta_0', \theta_1', \ldots, \theta_N') \) where \( \theta_0 = \theta_0' \), then

\[ \left( \frac{\theta_1}{1-\theta_0}, \ldots, \frac{\theta_N}{1-\theta_0} \right) >^m \left( \frac{\theta_1'}{1-\theta_0'}, \ldots, \frac{\theta_N'}{1-\theta_0'} \right) \]

\[ \iff \left( \theta_1, \ldots, \theta_N \right) >^m \left( \theta_1', \ldots, \theta_N' \right) \]
For example if $\theta = (0.90, 0.04, 0.03, 0.02, 0.01)$ and
$\theta' = (0.90, 0.03, 0.03, 0.03, 0.01)$, then $T_n, \theta \stackrel{\text{st}}{\geq} T_n, \theta'$.

**Remark 4.6.** Our results also allow us to make comparisons between systems with initially different numbers of bugs. Suppose initially system 1 is determined by $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$ and system 2 by $\theta' = (\theta_0', \theta_1', \ldots, \theta_N')$ where $\theta_i > 0$ and $\theta'_j > 0$ for all $i$ and $j$, and $N \leq N'$. We may view $\theta$ as an $N'+1$ dimensional vector by adding on $N'-N$ zero coordinates to form a new vector $\theta^*$. If $\theta^*$ and $\theta'$ satisfying conditions (i) and (ii) of Theorem 4.3, then

$$T_n, \theta \stackrel{\text{st}}{\geq} T_n, \theta' \quad \text{for any } n \leq N$$

and

$$B^m_\theta \leq_{\text{st}} B^m_\theta' \quad \text{for any } m.$$

In particular we may compare two systems which are initially both $\theta_0$ 100% free of bugs ($\theta_0 = \theta_0'$), and where system 1 has $N$ bugs with equal appearance probabilities ($\theta_i = (1-\theta_0)/N$ for $i = 1, \ldots, N$) while system 2 has $N'$ bugs with equal appearance probabilities ($\theta'_i = (1-\theta_0')/N'$ for $i = 1, \ldots, N'$). If $N \leq N'$ then

$$T_n, \theta \stackrel{\text{st}}{\geq} T_n, \theta' \quad \text{for any } n \leq N.$$

In addition, one has the following corollary:

**Corollary 4.7.** For a given $\theta = (\theta_0, \theta_1, \ldots, \theta_N)$, let
\[ \vec{\theta} = (\theta_0, 1-\theta_0, 0, \ldots, 0) \text{ and } \vec{\theta} = (\theta_0, \frac{1-\theta_0}{N}, \ldots, \frac{1-\theta_0}{N}). \text{ Then} \]

\[ T_n, \vec{\theta} \underset{st}{\geq} T_n, \vec{\theta} \underset{st}{\geq} T_n, \vec{\theta} \text{ for any } n \]

and \[ b^m_{\vec{\theta}} \overset{st}{\geq} b^m_{\vec{\theta}} \overset{st}{\geq} b^m_{\vec{\theta}} \text{ for any } m. \]

**Corollary 4.8.** Under the conditions of Theorem 4.3,

\[ E(T_n, \vec{\theta}) \geq E(T_n, \vec{\theta}'), \text{ for any } n \text{ and } E(B^m_{\vec{\theta}}) \leq E(B^m_{\vec{\theta}}') \text{ for any } m. \]

**Remark 4.9.** Suppose that \( \vec{\theta} = (\theta_0, \theta_1, \ldots, \theta_N) \) is the 'initial distribution' of our software system, and that a bug is eliminated from the system after it is encountered for the first \((k = 1)\) time. One may readily calculate the expected time to discovery of any \( n \) \((\leq N)\) bugs in the system in the following way. For \( 2 \leq k \leq N \), let \([i_1, \ldots, i_{k-1}]\) denote any ordered set of \( k-1 \) elements of \( \{1, 2, \ldots, N\} \) and let

\[ [b_{i_1}, \ldots, b_{i_{k-1}}] \] denote the event that bugs \( b_{i_1}, \ldots, b_{i_{k-1}} \) have been discovered in the system in the order \([i_1, \ldots, i_{k-1}]\).

We let \( \tau_1 = T_1, \vec{\theta} \) and \( \tau_k = T_k, \vec{\theta} - T_{k-1}, \vec{\theta} \) for \( k \geq 2 \). \( \tau_k \) is the random variable denoting the number of inputs \((\text{time})\) between the discovery of \( k-1 \) and \( k \) bugs in the system.

Now

\[ E(T_n, \vec{\theta}) = \sum_{k=1}^{n} E(\tau_k) \]

\[ = \frac{1}{(\theta_1 + \ldots + \theta_N)} + \sum_{k=2}^{n} \left[ \sum_{[i_1, \ldots, i_{k-1}]} \frac{E(\tau_k | [b_{i_1}, \ldots, b_{i_{k-1}}]) \cdot p([b_{i_1}, \ldots, b_{i_{k-1}}])}{p([b_{i_1}, \ldots, b_{i_{k-1}}])} \right]. \]
\[
\frac{1}{(\theta_1 + \ldots + \theta_N)} + \sum_{k=2}^{n} \left[ \frac{\theta_{i_1} \ldots \theta_{i_{k-1}}}{\sum_{j \neq 0, i_1, \ldots, i_{k-2}} \theta_j} \right] \\
\sum_{j \neq 0, i_1, \ldots, i_{k-1}} \theta_j \\
= \frac{1}{(\theta_1 + \ldots + \theta_N)} + \sum_{k=2}^{n} \left[ \frac{\theta_{i_1} \ldots \theta_{i_{k-1}}}{\sum_{j \neq 0, i_1} \theta_j} \right] \\
\sum_{j \neq 0, i_1, \ldots, i_{k-2}} \theta_j \\
\cdot \frac{\theta_{i_{k-1}}}{\sum_{j \neq 0, i_1, \ldots, i_{k-2}} \theta_j} \\
\sum_{j \neq 0, i_1, \ldots, i_{k-1}} \theta_j \\
\cdot \frac{1}{\sum_{j \neq 0, i_1, \ldots, i_{k-1}} \theta_j} \\
\]

For example, if \( N = 3 \) and \( n = 2 \), then the expected time to discovery of any 2 bugs is

\[
\frac{1}{\theta_1 + \theta_2 + \theta_3} \left( 1 + \frac{\theta_1}{\theta_2 + \theta_3} + \frac{\theta_2}{\theta_1 + \theta_3} + \frac{\theta_3}{\theta_1 + \theta_2} \right). \\
\]

The following table illustrates how the diversity of three bugs in a 90% fault free system influence the expected time to detection of any two bugs:

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( E(T_{2, \theta}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(.90, .04, .03, .03)</td>
<td>25.24</td>
</tr>
<tr>
<td>(.90, .05, .03, .02)</td>
<td>26.79</td>
</tr>
<tr>
<td>(.90, .06, .03, .01)</td>
<td>30.40</td>
</tr>
<tr>
<td>(.90, .08, .01, .01)</td>
<td>52.22</td>
</tr>
<tr>
<td>(.90, .08, .02, 0)</td>
<td>52.50</td>
</tr>
</tbody>
</table>

Figure 1.
§5. The Impact of Bug Diversity in the Jelinski-Moranda Model

In this section we consider the impact of fault diversity on bug detection for continuous models of the Jelinski-Moranda type. Let us assume that there are an unknown number \( N \) of faults in a software system, and that the appearance (or detection) times of the faults are independently and exponentially distributed random variables \( Z_i \) with failure rate \( \lambda_i \), \( i = 1, \ldots, N \). We assume that once a fault is detected it is eliminated from the system. Our model can hence be characterized by the vector \( (\lambda_1, \ldots, \lambda_N) \). We will use \( \bar{\lambda} \) to denote the average \( \frac{\sum_{i=1}^{N} \lambda_i}{N} \) of the individual fault detection rates. In the Jelinski-Moranda model, each bug or fault makes the same contribution to the overall failure rate, and \( \lambda_i = \bar{\lambda} \) for all \( i = 1, \ldots, N \).

Definition 5.1. For a given vector \( \lambda = (\lambda_1, \ldots, \lambda_N) \) of individual fault detection rates, the following random variables are of particular interest:

1. \( T_{n, \lambda} \) = time until any \( n \) bugs in the system are eliminated
2. \( B_{\lambda}^{t_0} \) = number of bugs or faults eliminated from the system during the time period \([0, t_0]\).

The following result is the analogue of Theorem 4.3 for the Jelinski-Moranda model.

Theorem 5.2. Let \( \lambda = (\lambda_1, \ldots, \lambda_N) \) and \( \lambda' = (\lambda'_1, \ldots, \lambda'_N) \) be two vectors of initial individual fault detection rates
for two software systems. If \( \lambda >^m \lambda' \), then

a) \( T_{n, \lambda} \overset{\text{st}}{\geq} T_{n, \lambda'} \) for any \( n \)

and b) \( B_{\lambda}^{t_0} \leq B_{\lambda'}^{t_0} \) for any \( t_0 \geq 0 \).

Proof. Let \( Z = (Z_1, \ldots, Z_N) \) (respectively \( Z' = (Z'_1, \ldots, Z'_N) \)) be independent exponentially distributed random variables with parameters \( \lambda = (\lambda_1, \ldots, \lambda_N) \) (respectively \( \lambda' = (\lambda'_1, \ldots, \lambda'_N) \)). Denote by \( Z(\cdot) \equiv (Z_1(\cdot), \ldots, Z_N(\cdot)) \) (respectively \( Z'(\cdot) = (Z'_1(\cdot), \ldots, Z'_N(\cdot)) \)) the vector of order statistics of \( Z \) (\( Z' \)).

Proschan and Sethuraman (1976) have shown that if \( \lambda \geq^m \lambda' \), then \( Z(\cdot) \overset{\text{st}}{\geq} Z'(\cdot) \), that is \( f(Z(\cdot)) \overset{\text{st}}{\geq} f(Z'(\cdot)) \) for all coordinate-wise increasing real valued functions \( f \).

a) \( f_1(w_1, \ldots, w_N) = w_n \) is coordinate-wise increasing and hence \( T_{n, \lambda} = f_1(Z(\cdot)) \overset{\text{st}}{\geq} f_1(Z'(\cdot)) = T_{n, \lambda'} \).

b) For any \( t_0 \geq 0 \), let \( \psi_{[0, t_0]} \) be the characteristic function of \( [0, t_0] \). Now \( f_2(w_1, \ldots, w_N) = \sum_{i=1}^{N} \psi_{[0, t_0]}(w_i) \) is coordinate wise decreasing and hence

\[
B_{\lambda}^{t_0} = f_2(Z(\cdot)) \overset{\text{st}}{\leq} f_2(Z'(\cdot)) = B_{\lambda'}^{t_0}.
\]
References


Proschan, F. and Sethuraman, J. (1976), Stochastic Comparisons of Order Statistics from Heterogeneous Populations, with Applications in Reliability, Journal of Multivariate Analysis, 6, 608-616.