Correlation in Multiversion Software

by

Toke Jayachandran

October 1996
This report was prepared in conjunction with research conducted for the Naval Postgraduate School and funded by the Naval Postgraduate School.

Reproduction of all or part of this report is authorized.

This report was prepared by:

Toke Jayachandran
Professor, Department of Mathematics

Reviewed by:

WALTER M. WOODS
Chairman

Released by:

DAVID W. NETZER
Dean of Research
**REPORT DOCUMENTATION PAGE**

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (OFR-0188), Washington, DC 20503.

<table>
<thead>
<tr>
<th>1. AGENCY USE ONLY (Leave Blank)</th>
<th>2. REPORT DATE</th>
<th>3. REPORT TYPE AND DATES COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>October 1996</td>
<td>Technical Report Jan - Sept. 96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE AND SUBTITLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CORRELATION IN MULTIVERSION SOFTWARE</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. AUTHOR(S)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TOKE JAYACHANDRA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</th>
<th>8. PERFORMING ORGANIZATION REPORT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naval Postgraduate School Monterey, CA 93943-5000</td>
<td>NPS-MA-003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</th>
<th>10. SPONSORING/MONITORING AGENCY REPORT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naval Postgraduate School Monterey, CA 93943</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. SUPPLEMENTARY NOTES</th>
<th>12a. DISTRIBUTION/AVAILABILITY STATEMENT</th>
<th>12b. DISTRIBUTION CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The views expressed in this report are those of the Author and do not reflect the official policy or position of the Department of Defense or the United States Government.</td>
<td>Approved for public release; distribution is unlimited.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. ABSTRACT (Maximum 200 words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>It has been established both theoretically [1] and experimentally [2], that independently developed redundant software versions fail dependently. Several probability models that account for this phenomenon of concurrent failures have appeared in the literature. Tomek et al., [3] proposed an intensity distribution that introduced a specific type of correlated failure pattern viz., pairwise correlation between software modules. They derived the intensity pmf for N = 2 and 3 modules and indicated the desirability of an efficient algorithm to compute the pmf for larger values of N. This paper contains an easily programmable algorithm to generate the pmf for any choice of N.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. SUBJECT TERMS</th>
<th>15. NUMBER OF PAGES</th>
<th>16. PRICE CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>redundant software, correlated failures, intensity distribution</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17. SECURITY CLASSIFICATION OF REPORT</th>
<th>18. SECURITY CLASSIFICATION OF THIS PAGE</th>
<th>19. SECURITY CLASSIFICATION OF ABSTRACT</th>
<th>20. LIMITATION OF ABSTRACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td>Unclassified</td>
<td>Unclassified</td>
<td></td>
</tr>
</tbody>
</table>

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89) Prescribed by ANSI Std 299-18
Correlation in Multiversion Software

Toke Jayachandran
Code MA/Jy
Naval Postgraduate School
Monterey, CA 93943

September, 1996

ABSTRACT

It has been established both theoretically [1] and experimentally [2], that independently developed redundant software versions fail dependently. Several probability models that account for this phenomenon of concurrent failures have appeared in the literature. Tomek et al., [3] proposed an intensity distribution that introduced a specific type of correlated failure pattern viz., pairwise correlation between software modules. They derived the intensity pmf for N = 2 and 3 modules and indicated the desirability of an efficient algorithm to compute the pmf for larger values of N. This paper contains an easily programmable algorithm to generate the pmf for any choice of N.

1 INTRODUCTION

The two principal techniques for software redundancy are N-version programming [4] and the recovery blocks [5]. Both require multiple independently developed software versions to achieve high reliability in software systems. Initially, it was believed that failures in independently produced software occur independently; and system reliability computations were based on this premise. It was subsequently demonstrated, both theoretically [1] and experimentally [2] that multiple versions can fail simultaneously for some choices of inputs. As a result, reliability estimates assuming independent failures can be overly optimistic. Several papers introducing probability models that allow for concurrent failures have appeared in the literature recently. Nicola and Goyal [6] proposed a model for simultaneous failure of independent software modules and the model has been shown to provide a good fit to the experimental data in [2]. Tomek et al., [3] introduced another model for generating the probability distribution (intensity distribution) of the number of modules (in an N-version system) that fail concurrently for a randomly selected input. The latter model incorporates the correlated failure syndrome into the intensity pmf through a parameter K that represents the probability that a pair of modules will produce identical outputs. They derived explicit
expressions for the pmf for \( N=2 \) and \( N=3 \) module software systems, and suggested that an efficient algorithm is needed to derive the pmf for larger values of \( N \). This paper presents such an algorithm for generating the intensity pmf for different choices of the parameters \( N \) and \( K \). The algorithm is easily programmable and is particularly suited for use with symbolic computation packages such as MAPLE\(^1\).

The Tomek et al., [3] model for correlated failure is described in Section II and the algorithm for deriving the intensity pmf for chosen values of \( N \) and \( K \) is presented in Section III. A MAPLE program for generating the pmf and the output of the program for \( N=5 \) and \( K=.1 \) are included in the Appendix.

2 A PROBABILITY MODEL FOR CORRELATED FAILURES

Consider a redundant software system with \( N \) independently developed modules. Let \( \Theta_N(X) \) be the proportion of modules (out of \( N \)) that fail (produce an incorrect output) for a randomly chosen input \( X \). Then \( \Theta_N(X) \) is a random variable assuming the values \( \{0, \ 1/N, \ 2/N, \ldots, \ 1\} \). \( \Theta_N(X) \) is called the intensity function and its probability distribution is referred to as the intensity distribution. For their probability based correlated failures model, Tomek et al. [3] assume that for each pair of modules, a proportion \( K \) of all possible inputs, will always generate identical outputs for the two modules. It is possible for two different pairs of modules to have identical inputs on two different sets of inputs, albeit the proportion of such inputs \( K \) is the same for all pairs. It is further assumed that a module will produce an incorrect output with probability \( p \). For \( N=2 \) modules, the space of all possible inputs is comprised of two subsets \( R \) and its complement \( R' \). \( R \) is the set of inputs for which the two modules will produce identical results, and for inputs from \( R' \) the module outputs are independent. The intensity function \( \Theta_2(X) \) assumes the values 0, 1/2, 1 and

\[
\begin{align*}
Pr[\Theta_2(X) = 0] &= Pr[X \epsilon R].Pr[\text{both module outputs are correct} | X \epsilon R] \\
&+ Pr[X \epsilon R'].Pr[\text{both module outputs are correct} | X \epsilon R'] \\
&= K(1 - p) + (1 - K)(1 - p)^2; \\
Pr[\Theta_2(X) = 1/2] &= Pr[X \epsilon R'].Pr[\text{exactly one output is correct} | X \epsilon R'] \\
&= 2(1 - K)p(1 - p); \\
Pr[\Theta_2(X) = 1] &= Pr[X \epsilon R].Pr[\text{both module outputs are correct} | X \epsilon R] \\
&+ Pr[X \epsilon R'].Pr[\text{both module outputs are correct} | X \epsilon R'] \\
&= Kp + (1 - K)p^2;
\end{align*}
\]

\(^1\text{MAPLE is a registered trademark of Waterloo Maple Software}\)
In the case of $N = 3$ modules, the input space is partitioned into 3 types of subsets $R_1$, $R_2$, and $R_3$ where $R_i$ ($i = 2, 3$) is the set of inputs for which exactly $i$ modules will produce identical results; $R_1$ is the set of inputs for which the module outputs are independent. There will be three subsets of the type $R_2$ and just one subset each of the types $R_1$ and $R_3$. The probabilities for the selection of an input from these subsets are $K^2$ for $R_3$, $3K(1 - K)$ for $R_2$ and $1 - K^2 - 3K(1 - K) = (1 - K)(1 - 2K)$ for $R_1$ and

$$Pr[\Theta_3(X) = j/3] = \Sigma_i Pr[X \in R_i].Pr[\text{exactly } j \text{ outputs are correct} \mid X \in R_i] \; j = 0 \ldots 3.$$

Therefore

$$\begin{cases} Pr[\Theta_3(X) = 0] = K^2(1 - p) + 3K(1 - K)(1 - p)^2 + (1 - K)(1 - 2K)(1 - p)^3; \\ Pr[\Theta_3(X) = 1/3] = K^2.0 + 3K(1 - K)p(1 - p)^2 + (1 - K)(1 - 2K)p(1 - p)^3; \\ Pr[\Theta_3(X) = 2/3] = K^2.0 + 3K(1 - K)p(1 - p) + (1 - K)(1 - 2K)p^2(1 - p); \\ Pr[\Theta_3(X) = 1] = K^2p + 3K(1 - K)p^2 + (1 - K)(1 - 2K)p^3. \end{cases} \tag{2}$$

The calculation of the probabilities of selecting an input from the subsets partitioning the input space, and the conditional pmf of $\Theta_N(X)$ becomes increasingly more difficult as the number of modules $N$ increases. An efficient algorithm that will perform the needed book keeping in a systematic fashion is presented in the next section.

3 AN ALGORITHM FOR GENERATING THE INTEENSITY DISTRIBUTION

For an $N$ module software system, the input space is partitioned in $N$ types of subsets $R_i$, $i = 1, 2, \ldots, N$. Inputs from subset type $R_i$ will result in identical outputs from $i$ of the $N$ modules. The number of subsets of type $R_i$, except for type $R_1$, is equal to $\binom{N}{i}$ the number of different ways of selecting $i$ modules from the available $N$ modules. There is just one subset of type $R_1$ and the module outputs are independent for inputs from this subset. The table below illustrates the pattern for the conditional probabilities $Pr[\Theta_N(X) = j/N \mid X \in R_i]$ when $N = 5$. 

3
TABLE 1

<table>
<thead>
<tr>
<th>(j/5 )</th>
<th>0</th>
<th>1/5</th>
<th>2/5</th>
<th>3/5</th>
<th>4/5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_5 )</td>
<td>((1 - p))</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(p)</td>
</tr>
<tr>
<td>( R_4 )</td>
<td>((1 - p)^2)</td>
<td>(p(1 - p))</td>
<td>0</td>
<td>0</td>
<td>(p(1 - p))</td>
<td>(p^2)</td>
</tr>
<tr>
<td>( R_3 )</td>
<td>((1 - p)^3)</td>
<td>(2p(1 - p)^2)</td>
<td>(p^2(1 - p))</td>
<td>(p(1 - p)^2)</td>
<td>(2p^2(1 - p))</td>
<td>(p^3)</td>
</tr>
<tr>
<td>( R_2 )</td>
<td>((1 - p)^4)</td>
<td>(3p(1 - p)^3)</td>
<td>(3p^2(1 - p)^2 + p(1 - p)^3)</td>
<td>(p^3(1 - p) + 3p^2(1 - p)^2)</td>
<td>(3p^3(1 - p))</td>
<td>(p^4)</td>
</tr>
<tr>
<td>( R_1 )</td>
<td>((1 - p)^5)</td>
<td>(5p(1 - p)^4)</td>
<td>(10p(1 - p))</td>
<td>(10p^3(1 - p)^2)</td>
<td>(5p^4(1 - p))</td>
<td>(p^5)</td>
</tr>
</tbody>
</table>

The probability entries in the table constitute a \(5 \times 6\) matrix \(P\) which can be expressed as the sum \(A + B\) of the two triangular matrices

\[
\begin{bmatrix}
(1 - p) & 0 & 0 & 0 & 0 & 0 \\
(1 - p)^2 & p(1 - p) & 0 & 0 & 0 & 0 \\
(1 - p)^3 & 2p(1 - p)^2 & p^2(1 - p) & 0 & 0 & 0 \\
(1 - p)^4 & 3p(1 - p)^3 & 3p^2(1 - p)^2 & p^3(1 - p) & 0 & 0 \\
(1 - p)^5 & 5p(1 - p)^4 & 10p^2(1 - p)^3 & 10p^3(1 - p)^2 & 5p^4(1 - p) & p^5
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & p \\
0 & 0 & 0 & 0 & p(1 - p) & p^2 \\
0 & 0 & 0 & p(1 - p)^2 & 2p^2(1 - p) & p^3 \\
0 & 0 & p(1 - p)^3 & 3p^2(1 - p)^2 & 3p^3(1 - p) & p^4 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

The above pattern persists for all \(N\) and the two \(N \times N + 1\) matrices, in the general case, have the form \(A = (a_{ij})\) and \(B = (b_{ij})\) where

\[
A = \begin{bmatrix}
(1 - p) & 0 & 0 & \ldots & 0 & p \\
(1 - p)^2 & p(1 - p) & 0 & \ldots & 0 & p^2 \\
(1 - p)^3 & 2p(1 - p)^2 & p^2(1 - p) & \ldots & 0 & p^3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(1 - p)^N & (N-1)p(1 - p)^{N-2} & (N-2)p^2(1 - p)^{N-3} & \ldots & p^{N-1}(1 - p) & p^N
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 & p \\
0 & 0 & 0 & \ldots & 0 & p^2 \\
0 & 0 & p(1 - p) & \ldots & 0 & p^3 \\
0 & 0 & p(1 - p)^2 & \ldots & 0 & p^4 \\
0 & 0 & 0 & \ldots & 0 & 0
\end{bmatrix}
\]
\[a_{ij} = \begin{cases} 
\binom{i-1}{j-1} p^{(j-1)} (1-p)^{(i-j+1)} & \text{for } j > i = 1, 2, \ldots, N-1 \\
\binom{N}{j-1} p^{(j-1)} (1-p)^{(N-j+1)} & \text{for } i = N \\
0 & \text{otherwise} 
\end{cases} \] (4)

\[b_{ij} = \begin{cases} 
\binom{i-1}{N+1-j} p^{i+j-N-1} (1-p)^{N+1-j} & \text{for } j > N - i = 1, 2, \ldots, N-1 \\
0 & \text{otherwise} 
\end{cases} \] (5)

The entries in the matrix \(P = (a_{ij} + b_{ij})\) are the conditional probabilities \(Pr[\Theta_N = (j-1)/N | X \in R_i], i = 1, \ldots, N\) and \(j = 1, \ldots, N + 1\). The unconditional probabilities or the intensity distribution is obtained by multiplying the matrix \(P\) on the left by the 1-row matrix \(Q = [q_1, q_2, \ldots, q_n]\) where

\[q_i = \binom{N}{i} K^{N-i-1}(1-K)^i \text{ for } i = 1, 2, \ldots, N-1\] (6)

\[q_N = 1 - \sum q_i \] (7)

Note that \(Q\) is just the vector of probabilities for an input to be in each of the subset types \(R_N, R_{N-1}, \ldots, R_1\).

The algorithm for computing the intensity pmf of \(\Theta_N(X)\) can be described by the following 3-step process. For specified values of the parameters \(N\) and \(K\)

1. Determine the 1-row matrix \(Q\) of input probabilities.

2. Evaluate the matrix \(P = A + B\), the matrix of conditional probabilities, 
   \(Pr[\Theta_N = (j-1)/N | X \in R_i]\).

3. Compute the matrix product \(Q \times P\) which is a \(1 \times N + 1\) matrix to obtain the intensity pmf of \(\Theta_N(X)\).
APPENDIX

Three MAPLE procedures \( r(N,K) \), \( Q(N,K) \) and \( P(N,K) \) to generate \( q_N \) the matrix \( Q \) in (6) and (7) and the matrix \( P = A + B \) in (3) are shown below. A printout of the MAPLE output creating these procedures and the computational results for \( N = 5 \) and \( K = .1 \) is also included.

\[
r: = (N,K) \rightarrow 1 - \text{sum}((\text{binomial}(n,i))^*(K^{N-1-i})*(1-K)^i), i = 0..N-2);
\]

\[
Q: = (N,K) \rightarrow \text{array}([\text{seq(binomial}(n,i)*K^{N-i-1}*(1-K)^i, i = 0..n-2), r(n,k)])};
\]

\[
P: = \text{proc}(N,K) \text{ local A,B,C,s,t;}
A: = \text{array}(1..N,1..N+1):
\text{for s to N do}
\text{for t to N + 1 do}
\text{if s < N then}
\text{if s > t - 1 then}
A[s,t] = (\text{binomial}(s-1, t-1)^*(K^{t-1})*(1-K)^{(s-t-1)})
\text{else A[s,t]} = 0 \text{fi;}
\text{else A[s,t]} = (\text{binomial}(n,t-1)^*(K^{t-1})*(1-K)^{(N-t-1)}); fi; \text{ od;}
\text{od;}
B: = \text{array}(1..N,1..N + 1):
\text{for s to N do}
\text{for t to N + 1 do}
\text{if s < N then}
\text{if t > N - s + 1 then}
B[s,t] = (\text{binomial}(s-1,N+1-t)^*K^{s+t-N-1}*((1-K)^{(N+1-t)}))
\text{else B[s,t]} = 0 \text{fi;}
\text{else B[s,t]} = 0 \text{fi; od;}
\text{od;}
C: = \text{evalm}(A + B);
\text{end;}
\]

Finally, the MAPLE expression

\[
\text{evalm}(Q(N,K) &* P(N,K)); \quad \text{will display the desired intensity pmf.}
\]
\[ r := (N, K) \rightarrow 1 - \sum_{i=0}^{N-2} \binom{N}{i} K^{N-1-i} (1-K)^i \]

\[ Q := (N, K) \rightarrow \text{array}([\{\text{seq}(\binom{N}{i}, i=0..N-2), r(N, K)\}] ) \]

\[ P := \text{proc}(N, K) \text{ local A, B, C, s, t; } \]
\[ A := \text{array}(1..N, 1..N+1); \]
\[ \text{for } s \text{ to } N \text{ do } \]
\[ \text{for } t \text{ to } N+1 \text{ do } \]
\[ \text{if } s < N \text{ then } \]
\[ \text{if } s > t-1 \text{ then } \]
\[ A[s, t] := (\binom{s-1}{t-1}) K^{t-1} (1-K)^{s-t+1} \]
\[ \text{else } A[s, t] := 0 \text{ fi; } \]
\[ \text{else } A[s, t] := (\binom{N}{t-1}) K^{t-1} (1-K)^{N-t+1} \text{ fi; } \]
\[ \text{od; } \]
\[ B := \text{array}(1..N, 1..N+1); \]
\[ \text{for } s \text{ to } N \text{ do } \]
\[ \text{for } t \text{ to } N+1 \text{ do } \]
\[ \text{if } s < N \text{ then } \]
\[ \text{if } t-1 < s \text{ then } \]
\[ A[s, t] := \binom{s-1}{t-1} K^{t-1} (1-K)^{s-t+1} \]
\[ \text{else } A[s, t] := 0 \text{ fi; } \]
\[ \text{else } A[s, t] := \binom{N}{t-1} K^{t-1} (1-K)^{N-t+1} \text{ fi; } \]
\[ \text{od; } \]
\[ P := \text{proc}(N, K) \text{ local A, B, C, s, t; } \]
\[ A := \text{array}(1..N, 1..N+1); \]
\[ \text{for } s \text{ to } N \text{ do } \text{ for } t \text{ to } N+1 \text{ do } \]
\[ \text{if } s < N \text{ then } \]
\[ \text{if } t-1 < s \text{ then } \]
\[ A[s, t] := \binom{s-1}{t-1} K^{t-1} (1-K)^{s-t+1} \]
\[ \text{else } A[s, t] := 0 \text{ fi; } \]
\[ \text{else } A[s, t] := \binom{N}{t-1} K^{t-1} (1-K)^{N-t+1} \text{ fi; } \]
\[ \text{od; } \]
\[ B := \text{array}(1..N, 1..N+1); \]
\[ \text{for } s \text{ to } N \text{ do } \text{ for } t \text{ to } N+1 \text{ do } \]
\[ \text{if } s < N \text{ then } \]
\[ \text{if } N-s+1 < t \text{ then } B[s, t] := \binom{s-1}{N-t+1} K^{t-1} (1-K)^{N-t+1} \]
\[ \text{else } B[s, t] := 0 \text{ fi; } \]
else \( B[s, t] := 0 \)
fi
od;
C := evalm(A + B)
end

> \( r(5, .1) \);

.1854000000

> \( Q(5, .1) \);

\[ [.0001, .0045, .0810, .7290, .1854000000] \]

> \( P(5, .1) \);

\[
\begin{bmatrix}
.9 & 0 & 0 & 0 & 0 & .1 \\
.81 & .09 & 0 & 0 & .09 & .01 \\
.729 & .162 & .009 & .081 & .018 & .001 \\
.6561 & .2187 & .0972 & .0252 & .0027 & .0001 \\
.59049 & .32805 & .07290 & .00810 & .00045 & .00001 \\
\end{bmatrix}
\]

> evalm(Q(5, .1) \&* P(5, .1));

\[ [.6505577460, .2337797770, .0851034600, .02643354000, .003914730000, .0002107540000] \]

>
REFERENCES


Distribution List

Director
Defense Technology Information Center
Cameron Station
Alexandria, VA 22314

Research Office
Code 81
Naval Postgraduate School
Monterey, CA 93943

Library
Code 52
Naval Postgraduate School
Monterey, CA 93943

Professor Toke Jayachandran
Department of Mathematics
Naval Postgraduate School
Monterey, CA 93943

Chairman, W. Max Woods
Department of Mathematics
Naval Postgraduate School
Monterey, CA 93943

No. of copies
2
1
2
5
2