

Reliability Bounds for Fault-Tolerant Systems With Competing Responses to Component Failures

Larry D. Lee

NASA Langley Research Center, Hampton

Key Words—Fault-tolerant system, Semi-Markov model, Reliability bound, Competing system-response.

Reader Aids—

Purpose: Present & analyze a model

Special math needed for explanations: Probability theory

Special math needed to use results: Same

Results useful to: Analysts of fault-tolerant systems

Abstract—Bounds are established for the probability of failure of fault-tolerant systems. The underlying failure and recovery process is assumed to follow a semi-Markov model in which the potential sojourn times of component failures have exponential distributions and those of system responses have general distributions. A product form of the bounds is derived from a model which provides for competing responses to component failures. Bounds are calculated in terms of integral factors which depend on component failure rates and the actual distributions of response times. Bounds are also calculated in terms of percentiles, conditional mean response times, and certain transition probabilities. Besides providing a computationally efficient method of estimating reliability from a flexible model, the bounds seem reasonably tight for a fairly wide range of cases. An example is given to illustrate calculating the bounds for a model with competing system responses.

1. INTRODUCTION

Background

Fault-tolerant hardware and software refer to a methodology for structuring systems to achieve high reliability. In critical applications such as the safe control of aircraft, special techniques [1-3] are used to detect and isolate failed components, switch in spare components, and switch out failed components. The underlying failure and recovery process can often be represented by a time-continuous Markov or semi-Markov model in which the states of the model describe the system status in terms of the numbers of operational components, numbers of active components, available spare components, etc. The jumps between states correspond to the occurrence of component failures or to system responses to previously failed components. The time parameter of the process can be the elapsed time since the start of a mission or local time since a particular event.

Within this framework it is usually of interest to estimate reliability, possibly for the purpose of studying the effect on reliability of varying the number of com-

ponents, their configuration, or other system design parameters which affect reliability. For example, the choice of decision rules [2] concerning the number of errors to be detected in the output of a component before the component is deactivated can affect system response times, hence affect reliability. Except for simple models which assume the failure and response times are exponentially distributed, estimates of reliability involve computationally difficult, multivariate integrals. Considerable research has been devoted to estimating reliability on the basis of various models: Stiffler, Bryant, Guccione [4], Ng & Avizienis [5], and other models and techniques as surveyed in Geist & Trivedi [6].

Description

The main purpose of this paper is to derive and present several forms of bounds for unreliability which generalize the bounds suggested by White [7] to a model that provides for competing responses to component failures. Although restricted to single system responses (ie, noncompeting responses), White's bounds are derived from a flexible semi-Markov model in which component failure times have exponential distributions and response times have general distributions. His bounds have a product form with factors that depend on component failure rates, and means and standard deviations of response times. In the present formulation, competing responses can correspond to the process of deactivating failed components and activating spare components; eg, deactivating component *A* vs component *B*. Rather than standard deviations of response times, the bounds are given in terms of integral factors which depend on the actual distributions of response times. Other forms of the bounds are given in terms of fewer parameters: percentiles, conditional mean response holding times, and certain transition probabilities.

Some advantages of a method based on bounds are: 1)it gives a computationally efficient technique for estimating reliability, 2)estimates can be given from experimentally derived information without making restrictive assumptions concerning the form of the Cdf's of response times, 3)a user of the bounds always knows whether he is under or overestimating the true unreliability (assuming the model is correct), 4)an estimate of the error is available whenever the bounds are calculated.

Sections 2 and 3 describe the model, section 4 derives the general form of the bounds, and sections 5 and 6 give bounds that require less detailed information concerning the Cdf's of response times. In section 7 an example illustrates calculating the bounds for a model with competing

system responses. Section 8, the final section, discusses tightness of the bounds.

2. NOTATION, NOMENCLATURE, AND ASSUMPTIONS

Notation

Λ	a set, $\{1, 2, \dots, k\}$, of possible system states; these are mapped 1-1 to a set of physical system states, eg, a physical state might be the number of operational components, status of system response, and current system configuration
$R(i)$	subset of Λ consisting of states the system can enter from state i as a result of system responses to previously failed components
$\bar{R}(i)$	states the system can enter from state i as a result of component failures
$T(i, j)$	potential sojourn time from state i to state j , a r.v.
$G(x; i, j)$	Cdf of $T(i, j)$, $j \in R(i)$; ie, Cdf of a system response time
$\bar{G}(x; i, j)$	Sf, $1 - G(x; i, j)$
$W(i)$	holding time in state i
z_0, z_1, \dots, z_n	any series of states that terminates in an absorbing state, z_n
Z_0, Z_1, \dots, Z_n	r.v.'s for which z_0, z_1, \dots, z_n are realizations
$\theta(i)$	$\Pr\{Z_0 = i\}$
$\theta(i, j)$	$\Pr\{Z_{m+1} = j Z_m = i\}$, $i, j \in \Lambda$, $m = 1, 2, \dots$

Other, standard notation is given in "Information for Readers & Authors" at rear of each issue.

Nomenclature

Potential sojourn time: Elapsed time, measured from the instant of entering a state, until the system enters another state

Sojourn (holding) time: Time, measured from the instant of entering a state, until the system changes state; more precisely, the smallest of the potential sojourn times

Assumptions

1. The failure and recovery process is semi-Markov; that is, the potential sojourn times are mutually statistically independent.

2. The potential sojourn times of component failures have exponential distributions.

3. System response times have general, continuous distributions.

4. The state space Λ has a finite number of elements and the system cannot re-enter a state; models with an infinite number of loops between states are not discussed.

5. All states that a system can enter from a particular state are known and states that represent system failure (absorbing states) are also known.

3. REMARKS CONCERNING THE MODEL

The model providing a basis for estimating reliability is that of the usual semi-Markov process (Heyman & Sobel, [8, p 196], Cox & Isham, [9, p 55]) in which the sequence of states entered over time is determined by a Markov chain and attached to each pair of states is a r.v. $T(i, j)$ representing a potential sojourn time from state i to state j , all potential sojourn times being mutually s -independent. If $T(i, j)$ corresponds to a state change resulting from a component failure, its Cdf is that of an exponential distribution with rate parameter $\lambda(i, j)$. Otherwise, $T(i, j)$ represents a system response time and has Cdf $G(x; i, j)$.

To calculate unreliability, we must consider each series of states z_0, z_1, \dots, z_n a system can enter over time and which terminates in an absorbing state. Typically, z_0 represents an initial state wherein all components are operational. The system enters some state z_1 as the result of a component failure, enters z_2 as a result of another component failure or, as a result of a system response to a previously failed component, and so on. Because the sojourn times $W(z_0), W(z_1), \dots, W(z_{n-1})$ between state changes are conditionally (given a particular series of states) s -independent, and Z_0, Z_1, \dots, Z_n follow a Markov chain [8, 9] their joint pdf is:

$$\theta(z_0) \prod_{i=1}^{n-1} \theta(z_i, z_{i+1}) dQ(w_i; z_i, z_{i+1}) \quad (1)$$

$\theta(i, j)$ transition probabilities

$$Q(x; z_m, z_{m+1}) = \Pr\{W(z_m) \leq x | Z_m = z_m, Z_{m+1} = z_{m+1}\}.$$

To interpret (1) in terms of the basic r.v.'s $\{T(i, \ell)\}$ suppose the system enters state $z_m = i$ after m state changes. Then $W(i)$ and Z_{m+1} can be represented as $W(i) = \min\{T(i, \ell) : \ell \in R(i) \cup \bar{R}(i)\}$ and $Z_{m+1} = j$ if and only if $T(i, j) = W(i)$; that is, the system enters state j from state i if and only if $T(i, j)$ is smaller than all other potential sojourn times attached to state i . It follows that:

$$\theta(i, j) Q(x; i, j) = \Pr\{W(i) \leq x, T(i, j) = W(i) | Z_m = i\}.$$

Each factor $\theta(i, j) dQ(x; i, j)$ of (1) can be calculated in terms of the Cdf's of the $\{T(i, \ell)\}$ from assumptions 1-3:

$$\lambda(i, j) \exp(-\lambda_i x) \prod_{\ell \in \bar{R}(i)} \bar{G}(x; i, \ell) dx, j \in \bar{R}(i) \\ \exp(-\lambda_i x) \prod_{\ell \in R(i)} \bar{G}(x; i, \ell) dG(x; i, j), j \in R(i) \quad (2)$$

$$\lambda_i \equiv \sum_{\ell \in \bar{R}(i)} \lambda(i, \ell)$$

the products in (2) involve only the indices of response times.

Results in [11, 12] show that if we begin with a model for which the response times associated with a particular state are s -dependent, and if the $Q(x; i, j)$ are continuous Cdfs, then there exists mutually s -independent r.v.'s, $\{T(i, \ell)\}$, having Cdf's that satisfy (2). This suggests that although it is unnecessary to assume that competing response times are s -independent, little generality is added by weakening assumption 1.

4. GENERAL FORM OF THE BOUNDS

Let A, B, C partition the indices of the transient states, $z_0, z_1, z_2, \dots, z_{n-1}$, in the following way:

$$A \equiv \{i: R(z_i) = \phi, i = 0, 1, 2, \dots, n-1\}$$

$$B \equiv \{i: R(z_i) \neq \phi, z_{i+1} \in R(z_i), i = 0, 1, 2, \dots, n-1\}$$

$$C \equiv \{i: R(z_i) \neq \phi, z_{i+1} \in \bar{R}(z_i), i = 0, 1, 2, \dots, n-1\}$$

That is, each state, $z_i, i = 0, 1, 2, \dots, n-1$ can be classified according to whether: all states which can be reached (in one step) from z_i are those that result from component failures; at the next state, a system response pre-empts component failures; at the next state, a component failure pre-empts system responses.

The main objective here is to give bounds for the probability, P say, of entering z_0, z_1, \dots, z_n by time t . The method of deriving the bounds is motivated by the assumption that, for highly reliable systems, component failure times are quite large and response times are quite short. From the definitions of A, B, C , the $W(z_i), i \in B \cup C$ are minima of sets of r.v.'s for which at least one member is a response time. Thus, $W(z_i)$ can be no larger than any response time attached to z_i . Let $T \equiv W(z_0) + W(z_1) + \dots + W(z_{n-1})$ be the elapsed time since the system occupies z_0 until entering z_n . The upper bound for P is calculated by excluding the potentially smaller sojourn times from T to approximate T by $\sum_A W(z_i)$. The lower bound is calculated by replacing each $W(z_i), i \in B \cup C$, in T by certain constants.

Define

$$P \equiv \Pr\{T \leq t, Z_0 = z_0, Z_1 = z_1, \dots, Z_n = z_n\}$$

$$P_U \equiv \Pr\{\sum_A W(z_i) \leq t, Z_0 = z_0, \dots, Z_n = z_n\}$$

$$P_L \equiv \Pr\left\{\sum_A W(z_i) \leq t - \Delta, W(z_i) \leq \delta_i, i \in B \cup C, Z_0 = z_0, \dots, Z_n = z_n\right\}$$

$$S \equiv \{(w_0, w_1, \dots, w_{n-1}): w_0 + w_1 + \dots + w_{n-1} \leq t\}$$

$$S_U \equiv \{(w_0, w_1, \dots, w_{n-1}): \sum_A w_i \leq t\}$$

$$S_L \equiv \{(w_0, w_1, \dots, w_{n-1}): \sum_A w_i \leq t - \Delta, w_i \leq \delta_i, i \in B \cup C\}$$

$$\Delta \equiv \sum_{B \cup C} \delta_i \quad (3)$$

where $\delta_i, i \in B \cup C$ are arbitrarily chosen nonnegative constants. The inequality, $P_L \leq P \leq P_U$, holds because $S_L \subseteq S \subseteq S_U$.

From the definitions of A, B, C , and from (2) the joint pdf of $W(z_0), W(z_1), \dots, W(z_{n-1}), Z_0, Z_1, \dots, Z_n$ is:

$$\begin{aligned} & \theta(z_0) \prod_A \{\lambda(z_i, z_{i+1})/\lambda_i\} \lambda_i \exp(-\lambda_i w_i) \\ & \prod_B \{\exp(-\lambda_i w_i) \prod_{\ell \neq z_i} \bar{G}(w_i; z_i, \ell) dG(w_i; z_i, z_{i+1})\} \\ & \prod_C \{\lambda(z_i, z_{i+1}) \exp(-\lambda_i w_i) \prod_{\ell} \bar{G}(w_i; z_i, \ell) dw_i\} \end{aligned} \quad (4)$$

By integrating this pdf over the regions defined by S_L and S_U , we get:

$$P_U = \theta(z_0) H(t) \prod_A a_i \prod_B b_i \prod_C c_i \quad (5)$$

$$P_L = \theta(z_0) H(t - \Delta) \prod_A a_i \prod_B b'_i \prod_C c'_i \quad (6)$$

$$\lambda_i \equiv \sum_{\ell \in \bar{R}(z_i)} \lambda(z_i, \ell) \quad (7)$$

$$a_i \equiv \lambda(z_i, z_{i+1})/\lambda_i \quad (8)$$

$$b_i \equiv \int_0^\infty \exp(-\lambda_i y) \prod_{\ell \neq z_{i+1}} \bar{G}(y; z_i, \ell) dG(y; z_i, z_{i+1}) \quad (9)$$

$$c_i \equiv \int_0^\infty \exp(-\lambda_i y) \lambda(z_i, z_{i+1}) \prod_{\ell} \bar{G}(y; z_i, \ell) dy \quad (10)$$

$$b'_i \equiv \int_0^{\delta_i} \exp(-\lambda_i y) \prod_{\ell \neq z_{i+1}} \bar{G}(y; z_i, \ell) dG(y; z_i, z_{i+1}) \quad (11)$$

$$c'_i \equiv \int_0^{\delta_i} \exp(-\lambda_i y) \lambda(z_i, z_{i+1}) \prod_{\ell} \bar{G}(y; z_i, \ell) dy \quad (12)$$

$$H(x) \equiv \Pr\left\{\sum_A W(z_i) \leq x\right\} \quad (13)$$

The indices for each product shown in (9)-(12) vary only over the indices of response time distributions. The $W(z_i), i \in A$, in (13) are s -independent r.v.'s having exponential distributions with rate parameters $\lambda_i, i \in A$.

To calculate the upper bound for unreliability, (5) must be summed over all series z_0, z_1, \dots, z_n that terminate in an absorbing state. Similarly, (6) must be summed for all such series to give the lower bound for unreliability.

Before proceeding to discuss other forms of the bounds in sections 5 and 6, we discuss briefly how the choice of the δ_i can affect tightness of the bounds. By calculating $b_i - b'_i$ and $c_i - c'_i$ from (9)-(12), we get the

inequalities:

$$b_i - b'_i \leq \left\{ \prod_{\ell} \overline{G}(\delta_i; z_i, \ell) \right\} \int_{\delta_i}^{\infty} \exp(-\lambda_i y) dy$$

$$c_i - c'_i \leq \lambda(z_i, z_{i+1}) \left\{ \prod_{\ell} \overline{G}(\delta_i; z_i, \ell) \right\} \int_{\delta_i}^{\infty} \exp(-\lambda_i y) dy \quad (14)$$

where $\prod_{\ell} \overline{G}(\delta_i; z_i, \ell)$ is the probability that the smallest response time (at state z_i) exceeds δ_i .

Because $\prod_{\ell} \overline{G}(\delta_i; z_i, \ell)$ decreases as δ_i increases, $b_i - b'_i$ and $c_i - c'_i$ also decrease as δ_i increases. Although the effect of increasing the δ_i is to also decrease $H(t - \Delta)$ in (6), the calculations of section 8 suggest that the effect on tightness of decreasing $H(t - \Delta)$ can be quite minimal. As a general rule for the case of short response times, each δ_i should be chosen as an extreme percentile at the upper tail of the Cdf of the smallest response times.

5. BOUNDS BASED ON MEANS, PERCENTILES, AND TRANSITION PROBABILITIES

Notation and Definitions

T_i response holding time, $\min\{T(z_i, \ell): \ell \in R(z_i)\}$, in state z_i

$\overline{G}(x; z_i)$ the Sf, $\Pr\{T_i > x\} = \prod_{\ell} \overline{G}(x; z_i, \ell)$

$G(x; z_i)$ $1 - \overline{G}(x; z_i)$

μ_i mean of T_i

$\mu(\delta; z_i)$ conditional mean, $\{G(\delta; z_i)\}^{-1} \int_0^{\delta} y dG(y; z_i)$, of T_i given that $T_i < \delta$

$q(z_i, j)$ $\Pr\{T(z_i, j) = T_i\} = \int_0^{\infty} \prod_{\ell \neq j} \overline{G}(y; z_i, \ell) dG(y; z_i, j)$

$K(x; z_i, j)$ $\Pr\{T(z_i, j) \leq x | T(z_i, j) = T_i\} = \{q(z_i, j)\}^{-1} \int_0^x \prod_{\ell \neq j} \overline{G}(y; z_i, \ell) dG(y; z_i, j)$

The indices, ℓ , for the products in the definitions of $\overline{G}(x; z_i)$, $q(z_i, j)$, and $K(x; z_i, j)$ vary only over the set $R(z_i)$.

The goal of this and the next section is to suggest bounds that can be calculated from as few parameters as possible rather than the full detail of the Cdf's of response times, as needed in (9)-(12)

Consider the upper bound in (5). After some manipulation in which $\exp(-\lambda_i y)$ in (9) is replaced by 1, and the integral in (10) is written as a sum by integrating over $(0, \delta_i)$ and (δ_i, ∞) , a new upper bound is obtained by replacing b_i and c_i in (5) by:

$$b_{1i} \equiv q(z_i, z_{i+1}) \quad (15)$$

$$c_{1i} \equiv \lambda(z_i, z_{i+1}) \{ \eta(\delta_i; z_i) + \lambda_i^{-1} \overline{G}(\delta_i; z_i) \} \quad (16)$$

$$\eta(\delta; z_i) \equiv G(\delta; z_i) \mu(\delta; z_i) + \delta \overline{G}(\delta; z_i). \quad (17)$$

The information needed for (15)-(17) consists of the transition probabilities, $q(i, j)$, the conditional means, $\mu(\delta_i; z_i)$, and the probabilities, $G(\delta_i; z_i)$. If only a single response is competing with component failures at state z_i , then $b_{1i} = 1$. Since $G(\delta; z_i) \mu(\delta; z_i) \leq G(\delta; z_i) \delta$, the upper bound can be calculated without knowledge of the $\mu(\delta_i; z_i)$.

A new lower bound is obtained by replacing each of b'_i and c'_i in (6) by:

$$b'_{1i} \equiv \exp(-\lambda_i \delta_i) b_{1i} K(\delta_i; z_i, z_{i+1}) \quad (18)$$

$$c'_{1i} \equiv \lambda(z_i, z_{i+1}) \exp(-\lambda_i \delta_i) \eta(\delta_i; z_i). \quad (19)$$

If only a single response is competing with component failures at state z_i , then $K(\delta_i; z_i, z_{i+1}) = G(\delta_i; z_i)$; otherwise,

$$\sum_{\ell \in R(z_i)} q(z_i, \ell) K(\delta_i; z_i, \ell) = G(\delta_i; z_i)$$

$$\sum_{\ell \in R(z_i)} q(z_i, \ell) = 1, \quad (20)$$

and (15) and (18) involve a substantial number of parameters.

Remark 1. If little information concerning response times is available, a practical solution is to assume that system responses occur, with probability equal to 1, within intervals of known length, δ_i . Then

$$G(\delta_i; z_i) = 1, \mu_i = \eta(\delta_i; z_i), K(\delta_i; z_i, z_{i+1}) = 1 \quad (21)$$

and the information needed to calculate b_{1i} , c_{1i} , b'_{1i} , and c'_{1i} consists of the transition probabilities, $q(i, j)$, the upper limits, δ_i , and the means, μ_i .

Remark 2. Instead of (16), we could have replaced c_i by $\lambda(z_i, z_{i+1}) \mu_i$, which appears simpler than (16). However, $\eta(\delta_i; z_i)$ is needed in (19). Although estimates of μ_i and $\eta(\delta_i; z_i)$ might differ when calculated from experimental data, they would differ very little if δ_i is an extreme upper percentile of $G(x; z_i)$; μ_i and $\eta(\delta_i; z_i)$ are identical, as mentioned in (21), whenever δ_i is an upper limit of the support of $G(x; z_i)$.

6. SIMPLER BOUNDS

We now consider bounds that are derived from a class of distributions which is known in the reliability literature as distributions having proportional hazard rates. This class gives a flexible model for the Cdf's of response times, and substantially reduces the information needed to calculate the bounds. The proportional hazards model has been studied extensively [13] as a basis for nonparametric statistical methods.

To define this class, let $G_0(x)$ be a continuous baseline Cdf, and let C be the class of continuous Cdfs $G(x)$

generated from $G_0(x)$ in the following way:

$$\overline{G}(x) = \{\overline{G}_0(x)\}^\alpha, x \geq 0, \text{ for some } \alpha > 0 \quad (22)$$

Suppose the response time Cdf's at state z_i belong to \mathcal{C} for some baseline Cdf, $G_0(x)$. Then —

$$-\ln\{\overline{G}(x; z_i, \ell)\} = \alpha(z_i, \ell) \{-\ln\overline{G}(x)\},$$

for some constants, $\alpha(z_i, \ell) > 0, \ell \in R(z_i)$. (23)

Now replace each $G(y; z_i, \ell)$ in:

$$K(x; z_i, j) = \{\varrho(z_i, j)\}^{-1} \int_0^x \prod_{\ell \neq j} \overline{G}(y; z_i, \ell) dG(y; z_i, j)$$

by its representation in (23). Upon simplifying, we get:

$$K(x; z_i, j) = G(x; z_i) \quad (24)$$

where $G(x; z_i)$ is the response holding time Cdf.

This well known result [13, p 171] in the reliability literature, states that the response holding time T_i in state z_i is s -independent of the event $T(z_i, j) = T_i$.

To summarize, if we assume the response times have distributions with proportional hazard rates, the b'_{1i} in (18) are equal to:

$$b'_{2i} \equiv \exp(-\lambda_i \delta_i) b_{1i} G(\delta_i; z_i). \quad (25)$$

Bounds calculated from b_{1i} , c_{1i} , b'_{2i} , and c'_{1i} require information concerning only the $\varrho(z_i, \ell)$, $G(\delta_i; z_i)$, and $\mu(\delta_i; z_i)$.

7. EXAMPLE

The example illustrates a model with competing system responses. Although experimental data on single-system-responses is available in Lala & Smith [14], it is not used in this example other than as a guideline for what might be a reasonable assumption for a mean response time. The main purpose of this example is to illustrate the method of calculating the bounds.

Example Assumptions

1. A system has three active processor units and one spare.
2. The active units have failure rate λ .
3. The spare has failure rate γ .
4. The outputs of the active units are input to a voter and the system fails (gives incorrect output) if two or more active units fail.
5. If errors are detected in the output of any active unit, the system activates the spare and switches it with the failed unit.
6. The spare and one active unit function as a cooper-

ating pair; its members are switched at regular intervals to check the spare's operational status.

The desire to check the spare's operational status leaves open the possibility of switching in a failed spare at some instant when an active unit, other than the cooperating unit, has failed. As shown in figure 1, one of the active units, other than the cooperating unit, fails (state 1); the spare fails (state 2); the system, being unaware that either unit has failed, automatically switches the spare with the good active unit (state 3). State 3, as well as states 5 and 7, represent system failure since the system is not fault-tolerant at any instant when two of the three active units have failed. States 6 and 8 designate operational states that are attained when the system detects, identifies, and deactivates the failed active unit, and replaces it with the spare.

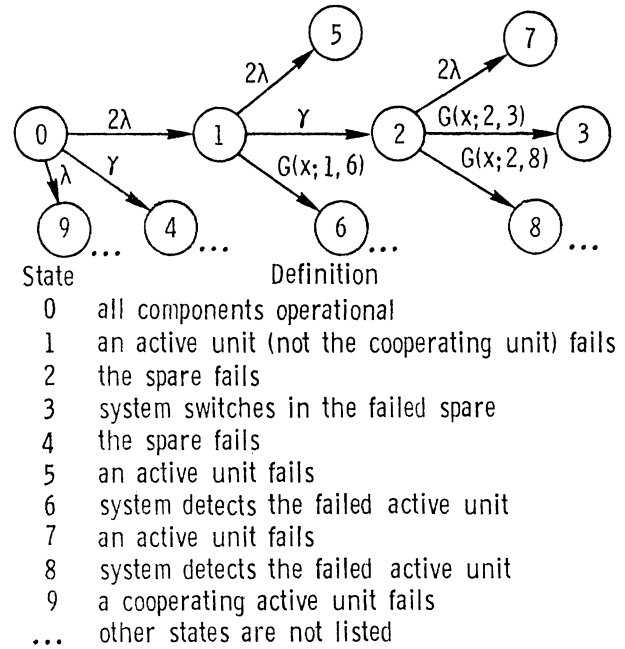


Fig. 1. State transition diagram for a system consisting of three active processor units and a spare.

The basic model for this example is defined by assumptions 1-5 in section 2 and by the Cdf's of the $\{T(i, \ell)\}$ for the state changes of figure 1. Exponentially distributed times of component failures are indicated by 2λ , γ , etc., in figure 1. Cdf's of response times are indicated by $G(x; i, j)$.

Consider calculating upper and lower bounds for the probability that the system enters states 0, 1, 2, 3 by time t . State 0 is the only state in the series 0, 1, 2 that involves only competing component failures, so $A = \{0\}$. Similarly, $C = \{1\}$ since the next state entered from state 1 is the result of a component failure that pre-empts a system response. In going from state 2 to 3, a system response pre-empts component failures, so $B = \{2\}$. The assumptions needed to calculate upper and lower bounds are those listed in 1-5 of section 2.

From (15)-(20) the bounds are,

$$P_U = H(t)a_0b_{12}c_{11}$$

$$P_L = H(t - \Delta)a_0b'_{12}c'_{11}$$

$$a_0 = 2\lambda(3\lambda + \gamma)^{-1}$$

$$b_{12} = \int_0^\infty \bar{G}(y; 2, 8)dG(y; 2, 3)$$

$$c_{11} = \gamma\{\eta(\delta_1; 1) + (2\lambda + \gamma)^{-1} \bar{G}(\delta_1; 1)\}$$

$$b'_{12} = b_{12} \exp(-2\lambda\delta_2) K(\delta_2; 2, 3)$$

$$c'_{11} = \gamma\eta(\delta_1; 1) \exp[-(2\lambda + \gamma)\delta_1]$$

$$\eta(\delta; 1) = G(\delta; 1) \mu(\delta; 1) + \delta \bar{G}(\delta; 1)$$

$$G(\delta; 2) = \varrho(2, 3) K(\delta; 2, 3) + \varrho(2, 8) K(\delta; 2, 8)$$

$$H(x) = 1 - \exp[-(3\lambda + \gamma)x].$$

Data (where appropriate, units are in hours):

$$\lambda = \gamma = 0.001; \delta_2 = \delta_3 = \delta = 0.04; t = 10.0;$$

$$\bar{G}(\delta; 1, 6) = 0; \mu(\delta; 1) = 0.02;$$

$$K(\delta; 2, 3) = K(\delta; 2, 8) = 0.9; \varrho(2, 8) = \varrho(2, 3) = 0.5$$

$$\text{Then } b_{12} = 0.5, \bar{G}(\delta; 1) = 0, G(\delta; 2) = 0.9, \eta(\delta; 1) = 0.02, \Delta = 0.08, P_U = 2.0 \times 10^{-7}, P_L = 1.8 \times 10^{-7}.$$

8. A STUDY OF TIGHTNESS

Although assumptions 1-5 (section 2) justify the upper and lower bounds of previous sections, they do not, however, ensure tightness. In this section, parameters affecting tightness are studied in terms of the relative error,

$$RE \equiv \{P_U - P_L\}/P_U. \quad (26)$$

For the less tight (than those of (9)-(12)) bounds given by (15)-(19),

$$RE = 1 - \{H(t - \Delta)/H(t)\} \prod_B (b'_{1i}/b_{1i}) \prod_C (c'_{1i}/c_{1i}) \quad (27)$$

$$b'_{1i}/b_{1i} = \exp(-\lambda_i \delta_i) K(\delta_i; z_i, z_{i+1})$$

$$c'_{1i}/c_{1i} = \exp(-\lambda_i \delta_i) \eta(\delta_i; z_i) \{\eta(\delta_i; z_i)$$

$$+ \lambda_i^{-1} \bar{G}(\delta_i; z_i)\}^{-1} \quad (28)$$

Since the number of parameters in (27) can be quite large, we consider only the case in which the response times

have distributions with support limited to bounded intervals; that is, it is assumed that:

$$\bar{G}(\delta_i; z_i, \ell) = 0, \ell \in R(z_i), i \in B \cup C. \quad (29)$$

Condition (29) implies $\bar{G}(\delta_i; z_i) = 0, i \in B \cup C$. This, in turn, implies:

$$K(\delta_i; z_i, z_{i+1}) = 1, i \in B \quad (30)$$

$$\mu_i = \eta(\delta_i; z_i), i \in C$$

$$b'_{1i}/b_{1i} = \exp(-\lambda_i \delta_i), i \in B$$

$$c'_{1i}/c_{1i} = \exp(-\lambda_i \delta_i), i \in C$$

$$RE = 1 - \{H(t - \Delta)/H(t)\} \exp(-\sum_{B \cup C} \lambda_i \delta_i) \quad (31)$$

RE in (31) does not depend upon the transition probabilities, $b_{1i} = \varrho(z_i, z_{i+1})$ or upon the mean response holding times, $\mu_i = \eta(\delta_i; z_i)$; thus a fairly wide range of cases can be studied by restricting attention to bounds calculated under (29).

Let m be the number of elements in A and let $\lambda_1, \lambda_2, \dots, \lambda_m$ be the rate parameters of the exponentially distributed r.v.'s, $W(z_i), i \in A$. Recursive calculation of $H(x)$ in (13) gives:

$$H_1(x) = 1 - \exp(-\lambda_1 x)$$

$$H_r(x) = \int_0^x H_{r-1}(x-y) \lambda_r \exp(-\lambda_r y) dy,$$

$r = 2, 3, \dots, m$. If $\lambda_1, \lambda_2, \dots, \lambda_m$ are all distinct, $H(x)$ can be calculated (Bartholomew, [15]) as a mixture:

$$H(x) = \sum_1^m \pi_i \{1 - \exp(-\lambda_i x)\}$$

of exponential Cdf's with weights π_i (possibly negative):

$$\pi_i \equiv \prod_{\substack{j=1 \\ j \neq i}}^m \lambda_j / (\lambda_j - \lambda_i) \quad i = 1, 2, \dots, m$$

Many different cases of (31) can be studied by varying $\lambda_i, i \in A \cup B \cup C$ and $\delta_i, i \in B \cup C$, over the interval $(0, \infty)$. Let q be the number of elements of $B \cup C$ and consider the following:

Example 1. $\lambda_i = (5 - i + 1)\lambda, i = 1, 2, \dots, m, m \leq 5,$

$$\lambda_i = \lambda, \delta_i = \delta, i \in B \cup C, q = 5.$$

Example 2. $\lambda_i = \lambda, i \in A \cup B \cup C, m = 1, 2, \dots, \delta_i = \delta,$

$$i \in B \cup C, q = 5.$$

For Example 1,

$$H(x) = 1 - \sum_{\ell=0}^{m-1} (5!/\ell!(5-\ell)!) [1 - \exp(-\lambda x)]^\ell \exp[-(5-\ell)\lambda x]$$

For Example 2,

$$H(x) = 1 - \exp(-\lambda x) \sum_{\ell=0}^{m-1} (\lambda x)^\ell / \ell!$$

All relative errors in table I were calculated for examples 1 and 2 and a mission time of $t = 10$ hours. The selection of ranges for λ (in failures per hour) and for δ (in hours) was arbitrary, although the Lala & Smith [14] data suggest that response times can be much smaller than the limits, δ , considered in table I. In examples 1 and 2, the $\Delta = 5\delta$ corresponds to a series of system states z_0, z_1, \dots, z_{n-1} for which the potential sojourn times attached to $q = 5$ states involve system responses and the remaining states, m of them, involve only component failures.

TABLE I
Relative Errors for Estimating the Probability of
Entering an Absorbing State: Examples 1 and 2

m	λ	δ	RE (Example 1)	RE (Example 2)
1	0.1000	0.02	0.0103	0.0157
1	0.1000	0.04	0.0205	0.0313
1	0.1000	0.06	0.0306	0.0467
1	0.0100	0.02	0.0087	0.0105
1	0.0100	0.04	0.0175	0.0210
1	0.0100	0.06	0.0262	0.0315
1	0.0010	0.02	0.0099	0.0100
1	0.0010	0.04	0.0197	0.0201
1	0.0010	0.06	0.0296	0.0301
1	0.0001	0.02	0.0100	0.0100
1	0.0001	0.04	0.0200	0.0200
1	0.0001	0.06	0.0300	0.0300
2	0.1000	0.02	0.0124	0.0237
2	0.1000	0.04	0.0248	0.0471
2	0.1000	0.06	0.0371	0.0701
2	0.0100	0.02	0.0181	0.0202
2	0.0100	0.04	0.0360	0.0403
2	0.0100	0.06	0.0538	0.0601
2	0.0010	0.02	0.0197	0.0199
2	0.0010	0.04	0.0392	0.0397
2	0.0010	0.06	0.0585	0.0592
2	0.0001	0.02	0.0199	0.0199
2	0.0001	0.04	0.0396	0.0396
2	0.0001	0.06	0.0590	0.0591

Table I shows that the relative error decreases as δ decreases but can increase as λ decreases, or as m increases. The increase in RE is limited, however, as λ decreases.

Consider the more general case in (31) and —

$$H(x) = \lambda_1 \lambda_2 \dots \lambda_m \int_D \exp\left(-\sum_A \lambda_i x_i\right) dx_1 dx_2 \dots dx_m$$

$$D \equiv \{(x_1, x_2, \dots, x_m): x_1 + x_2 + \dots + x_m \leq x\}.$$

Let the $\lambda_i, i \in A$, decrease to zero to get:

$$\lim H(t - \Delta)/H(t) = (1 - \Delta t^{-1})^m. \quad (32)$$

The limiting (as $\lambda_i \rightarrow 0, i \in A$) relative error is:

$$\lim RE = 1 - (1 - \Delta t^{-1})^m \exp\left(-\sum_{BC} \lambda_i \delta_i\right). \quad (33)$$

This limit is approached rapidly for intermediate calculations which are not shown in table I: for example 1 with $m = 2, \delta = 0.04, \Delta = 5\delta, t = 10.0$,

$$\begin{aligned} H(t - \Delta)/H(t) &= 0.96167 \text{ at } \lambda = 0.0100 \\ &= 0.96041 \text{ at } \lambda = 0.0001 \end{aligned}$$

whereas the limiting value from (32) is 0.96040. The limiting RE for this case is 0.0396 which is the same, up to four decimal places, as given in table I at $\lambda = 0.0001$.

These calculations suggest that the limiting RE is a good indicator of tightness. It embodies the main requirements for tightness; namely, short response times (small limits, δ_i), and large times to component failures (small λ_i). Further, it indicates that other parameters can affect tightness, such as the mission time t , and the numbers m and q of states in z_0, z_1, \dots, z_n that have indices in A and $B \cup C$, respectively. Except for the inconvenience of calculating $H(x)$, one can say the same thing for RE in (31).

ACKNOWLEDGMENT

I am grateful to the *Editor* and the referees for their comments and suggestions which led to the discussion of tightness in section 8. R. Butler of NASA Langley Research Center performed the calculations of table I. He has automated the calculation of the bounds given in an earlier version of this paper and the bounds given by A. White. An earlier version of this paper was a NASA technical paper.

REFERENCES

- [1] R. C. Montgomery, "Failure detection and control system reconfiguration: past, present, and future," *Systems Reliability Issues for Future Aircraft*, NASA CP-003. Available from: Langley Research Center; Hampton, VA USA, 1975, pp 69-78.
- [2] T. B. Smith, A. L. Hopkins, E. C. Hall, J. R. Howatt, J. H. Lala, "A fault-tolerant multiprocessor architecture for aircraft - vol II," NASA CR1-165911, (See [1]), 1977.

- [3] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol 12, 1976, pp 601-611.
- [4] J. J. Stiffler, L. A. Bryant, L. Guccione, "Care III final report, phase I," NASA CR-159122, (see [1]), 1974.
- [5] Y. W. Ng, A. Avizienis, "A model for transient and permanent fault recovery in closed fault-tolerant systems," in *Proc. 1976 Int. Symp. Fault-Tolerant Computer*, 1976 June.
- [6] R. M. Geist, K. S. Trivedi, "Ultrahigh reliability prediction in fault-tolerant computer systems," *IEEE Trans. Computers*, vol C32, 1980 Dec, pp 1118-1127.
- [7] A. White, "Upper and lower bounds for semi-Markov reliability models for reconfigurable systems," NASA CR-172340, 1984 April, (See [1]).
- [8] D. P. Heyman, M. J. Sobel, *Stochastic Models in Operations Research Volume III*, McGraw-Hill, 1984.
- [9] D. R. Cox, V. Isham, *Point Processes*, Chapman & Hall, 1980.
- [10] S. W. Lagakos, C. J. Sommer, M. Zelen, "Semi-Markov models for partially censored data," *Biometrika*, vol 65, No. 2, 1978, pp 311-317.
- [11] A. Tsiatis, "A nonidentifiability aspect of the problem of competing risks," *Proc. Nat. Acad. Sci. USA*, vol 72, 1975, pp 20-22.
- [12] D. R. Miller, "A note on independence of multivariate lifetimes in competing risk models," *The Annals of Statistics*, vol 5, 1977, pp 576-579.
- [13] J. D. Kalbfleisch, R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, 1980.
- [14] J. H. Lala, T. B. Smith, "Development and evaluation of fault-tolerant multiprocessor (FTMP) computer, vol III, FTMP test and evaluation," NASA CR-15336, 1983, see [1].
- [15] D. J. Bartholomew, "Sufficient conditions for a mixture of exponentials to be a probability density function," *Annals of Math. Stat.*, vol 40, 1969, pp 2183-2188.

AUTHOR

Larry D. Lee; System Validation Methods Branch; NASA Langley Research Center; Hampton, Virginia 23665 USA.

Larry D. Lee received the BS and MS degrees in mathematics at Illinois State University and the PhD in statistics from the University of Missouri at Columbia. His research interests are in reliability theory and statistical inference.

Manuscript TR84-128 received 1984 November 21; revised 1985 October 11. ★ ★ ★

EXTENDED ABSTRACT EXTENDED ABSTRACT EXTENDED ABSTRACT EXTENDED ABSTRACT EXTENDED ABSTRACT

Reliability Estimation in Stress-Strength Models: A Bayes Approach

M. Pandey

Banaras Hindu University, Varanasi

S. K. Upadhyay

Bararas Hindu University, Varanasi

Key Words—Weibull distribution, Scale and shape parameters, Stress-strength model, Bayes inference, Maximum likelihood estimation.

The paper provides a Bayes approach of drawing inference about the reliability of a 1-component system whose failure mechanism is simple stress-strength. The Bayes estimator of system reliability is obtained from data consisting of random samples from the stress and strength distributions, assuming each one is Weibull. The Bayes estimators of the four unknown shape and scale parameters of stress and strength distributions are also considered and these estimators are used in estimating the system reliability. The priors of the parameters of stress and strength distributions are assumed to be independent.

The Bayes credibility interval of the scale and shape parameters are derived using the joint posterior of the

parameters. The proposed procedure is illustrated by a numerical example based on generated data. The proposed Bayes approach gives point estimates and the posterior limits which are satisfactory. The complete paper is given in a separately available Supplement.

An earlier version of the complete paper was presented at the joint XIII Intern. Conf. on Stochastic Processes and Their Applications and the Ann. Conf. of the Indian Soc. for Theory of Probability and Its Applications, held 1983 Dec 17-21 at BHU in Varanasi, India.

SUPPLEMENT

- [1] NAPS document No. 04300-B; 10 pages in this supplement. For current ordering information, see "Information for Readers & Authors" in a current issue. Order NAPS document No. 04300, 95 pages. ASIS-NAPS; Microfiche Publications; POBox 3513, Grand Central Station; New York, NY 10163 USA.

AUTHORS

M. Pandey; Dept. of Zoology; Banaras Hindu University; Varanasi - 221 005 INDIA.

S. K. Upadhyay; Dept. of Statistics; Banaras Hindu University; Varanasi - 221 005 INDIA.

Manuscript TR83-159 received 1983 October 25; revised 1985 December 27. ★ ★ ★