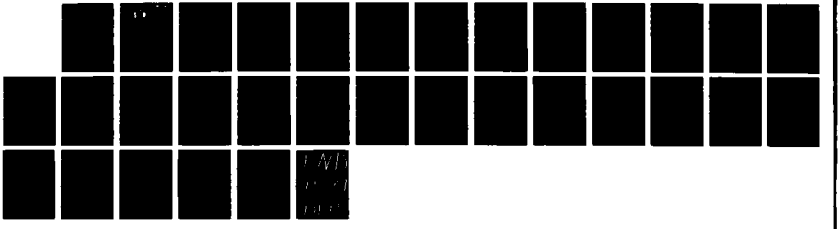
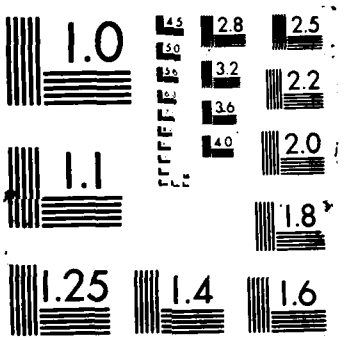
AD-A186 264 HOW ERRORS IN COMPONENT RELIABILITY AFFECT SYSTEM RELIABILITY..(U) NORTH CAROLINA UNIV AT CHAPEL HILL CURRICULUM IN OPERATIONS R.. G S FISHMAN JUL 87 1/1

UNCLASSIFIED UNC/ORSA/TR-87/3 AFOSR-TR-87-0994 F/G 12/3 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

SECU

AD-A186 264

DTIC FILE COPY

DOCUMENTATION PAGE

DTIC ELECTE OCT 19 1987 S D

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY: N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S):	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR. 87-0994
6a. NAME OF PERFORMING ORGANIZATION: University of North Carolina at Chapel Hill	6b. OFFICE SYMBOL (If applicable):
6c. ADDRESS (City, State and ZIP Code): 210 Smith Building 128A, Chapel Hill, NC 27514	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR	8b. OFFICE SYMBOL (If applicable): NM
8c. ADDRESS (City, State and ZIP Code): Building 410, Bolling AFB, Washington, DC 20332-6448	7b. ADDRESS (City, State and ZIP Code): Building 410, Bolling AFB, DC 20332-6448
	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-84-0140
	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 6.1102F; PROJECT NO.: 2304; TASK NO.: A5; WORK UNIT NO.:

11. TITLE (Include Security Classification): How Errors in Component Reliability Affect System Reliability

12. PERSONAL AUTHOR(S): George S. Fishman

13a. TYPE OF REPORT: Technical REPRINT	13b. TIME COVERED: FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day): July, 1987	15. PAGE COUNT: 28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Keywords: reliability, s-t connectedness

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

This paper studies how sampling variation in component reliability estimates affects the computation of system reliability that uses these estimates as input. Results show that relative bias in system reliability grows quadratically with the number of components for which each component reliability estimate is used, whereas the corresponding coefficient of variation grows linearly with this number of components. If these components are in parallel they lead to an understatement of system reliability. In series, they lead to an overstatement. The paper describes resampling schemes that eliminate bias without increasing the dominant variance term.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL: George S. Fishman	22b. TELEPHONE NUMBER (Include Area Code): 919-962-8401 — 22c. OFFICE SYMBOL: AFOSR/NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

OPERATIONS RESEARCH AND SYSTEMS ANALYSIS

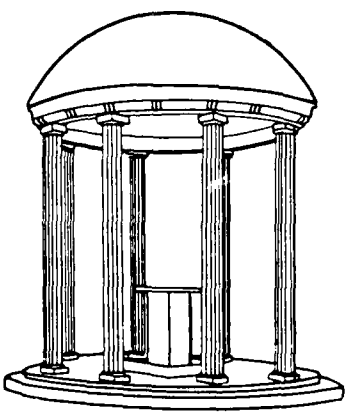
How Errors in Component Reliability Affect System Reliability

George S. Fishman

Technical Report No. UNC/ORSA/TR-87/3

July 1987

UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL



How Errors in Component Reliability Affect System Reliability

George S. Fishman

Technical Report No. UNC/ORSA/TR-87/3

July 1987

Department of Operations Research
University of North Carolina
Chapel Hill, North Carolina

This research was supported by the Air Force Office of Scientific Research under grant AFOSR-84-0140. Reproduction in whole or part is permitted for any purpose of the United States Government.

Abstract

This paper studies how sampling variation in component reliability estimates affects the computation of system reliability that uses these estimates as input. Results show that relative bias in system reliability grows quadratically with the number of components for which each component reliability estimate is used, whereas the corresponding coefficient of variation grows linearly with this number of components. If these components are in parallel they lead to an understatement of system reliability. In series, they lead to an overstatement. The paper describes resampling schemes that eliminate bias without increasing the dominant variance term.

Keywords: reliability, s-t connectedness

Introduction

Every computation of system reliability relies on the availability of numerical values for the reliabilities of components from which the system is constructed. If these numerical values were exact, then a direct computation of system reliability would at most suffer from numerical roundoff error. Since the numerical values of component reliabilities rarely are known with exactness, a system reliability computation customarily employs estimates of these quantities derived from test data. This substitution introduces an additional source of error attributable to the sampling variation inherent in the component reliability estimates. As the present paper shows, neglecting this source of error can produce a misleading system reliability.

This error manifests itself in bias and variance. For a system composed of several types of components where the system reliability computation uses a common component reliability estimate as input for all components of the same type, the relative bias in system reliability increases quadratically with each of the numbers of components of each type, whereas the corresponding coefficient of variation grows linearly with these numbers. For components of the same type in parallel, this system reliability computation understates true reliability. For components of the same type in series, the computation overstates reliability.

These results imply that for given component reliability estimates system reliability computations for two different

systems composed of exactly the same number of components of each type can have substantially different statistical error characteristics. While no method exists for reducing the variance of the system reliability base on component reliability estimates of fixed sample sizes, resampling schemes do allow one to eliminate bias without increasing the dominant variance term.

Section 1 introduces the notation for characterizing a system as a network. Section 2 gives the conventional estimator for system reliability and describes how one can use a confidence interval to assess its statistical accuracy. Section 3 shows how parallel and series systems affect statistical error and Section 4 extends the results to more general systems. Section 5 describes two resampling plans that eliminate bias while preserving the dominant variance term. Section 6 gives the conclusions of the study.

1. System Characteristics

As a basis for studying error, consider the network $G = (\underline{V}, \underline{E})$ with node set $\underline{V}$ and arc set $\underline{E}$. For convenience of exposition, assume that nodes represent components that function perfectly and that arcs represent components that fail randomly and independently. Hereafter, we treat the word component as synonymous with arc. To characterize G more completely, we define:

r = number of distinct types of components

p_i = probability that a component of type i functions $i=1,\ldots,r$

$\underset{\sim}{p} = (p_1,\ldots,p_r)$

$\underset{\sim}{E}_i$ = set of arcs that use components of type i $(\underset{\sim}{E}_i \cap \underset{\sim}{E}_j = \emptyset \quad i \neq j, \quad \underset{\sim}{E} = \bigcup_{i=1}^{r} \underset{\sim}{E}_i)$

$k_i = |\underset{\sim}{E}_i|$ number of components of type i

$\underset{\sim}{k} = (k_1,\ldots,k_r)$

e_{ij} = jth arc in $\underset{\sim}{E}_i$

$x_{ij} = 1$ if arc e_{ij} functions, $= 0$ otherwise

$x_i = \sum_{j=1}^{k_i} x_{ij}$ = number of arcs of type i that function

$\underset{\sim}{x} = (x_{11},\ldots,x_{1k_1}; x_{21},\ldots,x_{2k_2};\ldots;x_{r1},\ldots,x_{rk_r})$

$\underset{\sim}{X}$ = set of all arc states $\underset{\sim}{x}$

$$P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{p}) = \prod_{i=1}^{r} p_i^{x_i} (1-p_i)^{k_i - x_i} \qquad \underset{\sim}{x} \in \underset{\sim}{X}$$

= probability mass function of states in $\underset{\sim}{X}$

$\phi(\underset{\sim}{x}) = 1$ if the system functions, $= 0$ otherwise

$g(\underset{\sim}{p}) = \sum_{\underset{\sim}{x} \in \underset{\sim}{X}} \phi(\underset{\sim}{x})\, P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{p})$ = probability that the system functions.

We also assume that G describes a coherent system. A system of components is coherent if its structure $\{\phi(\underset{\sim}{x})\}$ is nondecreasing and each component is relevant. See Barlow and Proschan (1981, p. 6).

The system reliability $g(\underset{\sim}{p})$ can have diverse interpretations. For example, let $\underset{\sim}{T}$ denote a subset of $\underset{\sim}{V}$ and let

$\phi(\underset{\sim}{x}) = 1$ if all nodes in $\underset{\sim}{T}$ are connected when arc state $\underset{\sim}{x}$ occurs

$= 0$ otherwise.

Then $g(\underset{\sim}{p})$ denotes the probability that all nodes in T are connected. If $\underline{T} = \{s,t\}$, this is called the s-t connectedness problem. If $\underline{T} = \underline{V}$, it is called the all terminal connectedness problem.

Reliability in flow problems can also be characterized. Suppose that G is a directed acyclic flow network with source node s and terminal node t. Let

$$p_i = \text{pr (arc } j \text{ has flow capacity } b_i) \qquad b_i > 0$$

$$1 - p_i = \text{pr (arc } j \text{ has zero flow capacity)}$$

$$x_{ij} = 1 \quad \text{if arc } j \text{ in } \underline{E}_j \text{ has flow capacity } b_i$$

$$= 0 \quad \text{if arc flow capacity is zero}$$

and let

$$\phi(\underset{\sim}{x}) = \phi(\underset{\sim}{x},z) = 1 \quad \text{if the maximal s-t flow exceeds a specified demand z when state } \underset{\sim}{x} \text{ occurs}$$

$$= 0 \quad \text{otherwise.}$$

Then $g(\underset{\sim}{p}) = g(\underset{\sim}{p},z)$ denotes the probability that the maximal s-t flow in G exceeds z.

Although the exact computation of $g(\underset{\sim}{p})$ for these examples belongs to the NP-hard class of problems (Valiant 1979, Ball and Provan 1983, Provan 1986), we assume that for a particular network instance of interest, one can indeed effect the exact computation *if* $\underset{\sim}{p}$ *is known. If an exact computation proves infeasible and one resorts to the Monte Carlo method, then one needs to*

perform a more elaborate analysis to determine how statistical variation in the component reliability estimates interacts with the sampling variation that the Monte Carlo method induces.

2. Component Reliability Estimates

In practice, $\underset{\sim}{p}$ is not known exactly, but can be estimated from test data. Suppose one tests n_i components of type i for $i=1,\ldots,r$. Each test begins with a new component functioning. Let Z_{ij} denote the outcome of the jth test of component of type i where $Z_{ij} = 1$ if the component functions at the end of the test period and $Z_{ij} = 0$ if the component fails prior to the end of the test period. Presumably each component of type i is tested under identical conditions that resemble the system environment. Then one has the data vectors $\underset{\sim}{Z}_i = \{Z_{i1},\ldots,Z_{in_i}\}$ for $i=1,\ldots,r$ where the elements of $\underset{\sim}{Z}_i$ are independent and identically distributed with $p_i = EZ_{ij}$ $j=1,\ldots,n_i$, $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_r$ are independent, and

$$\hat{p}_i = n_i^{-1} \sum_{j=1}^{n_i} Z_{ij} \tag{1}$$

gives the maximum likelihood estimator of p_i with

$$E\hat{p}_i = p_i \tag{2}$$

$$\text{var } \hat{p}_i = p_i(1-p_i)/n_i \tag{3}$$

$$E(\hat{p}_i - p_i)^m = O(1/n_i^{\lfloor (m+1)/2 \rfloor}) \quad m=3,4,\ldots \quad \text{as} \quad n_i \to \infty \tag{4}$$

where $O(y)$ as $y \to L$ denotes a function f such that $\lim_{y \to L} |f(y)|/y$ is bounded. Observe from (2) that $\hat{p}_i$ is an unbiased estimator of p_i.

Let $\hat{\underset{\sim}{p}} = (\hat{p}_1, \ldots, \hat{p}_r)$. Then it is not unusual to estimate $g(\underset{\sim}{p})$ by $g(\hat{\underset{\sim}{p}})$. Although other methods exist for using test data to estimate component reliabilities, the appeal of the method that we adopt here arises from the well-understood sampling properties of $\hat{\underset{\sim}{p}}$, enabling us to concentrate on the statistical variation in $g(\hat{\underset{\sim}{p}})$ that substitution of $\hat{\underset{\sim}{p}}$ for $\underset{\sim}{p}$ in $g(\underset{\sim}{p})$ induces. As Gaver and Hoel (1970) show, other methods can lead to bias in component reliability estimates, which would force us to conduct a more complicated analysis to get at the sampling properties in the system reliability estimate.

As Sections 3 and 4 make clear, $g(\hat{\underset{\sim}{p}})$ generally either understates or overstates $g(\underset{\sim}{p})$ with regard to expectation. However, for the moment, we describe how one can globally assess the statistical accuracy of $g(\hat{\underset{\sim}{p}})$ based on confidence intervals computed for $\hat{p}_1, \ldots, \hat{p}_r$.

Let $Z_i = \sum_{j=1}^{n_i} Z_{ij}$. For each p_i we seek a $100 \times (1-\alpha)$ confidence interval $[p_i^*(Z_i, n_i),\ p_i^{**}(Z_i, n_i)]$

$$\mathrm{pr}[p_i^*(Z_i, n_i) \leq p_i \leq p_i^{**}(Z_i, n_i)] \geq 1 - \alpha.$$

Let

$$F_j(m, q) = \sum_{i=0}^{m} \binom{m}{i} q^i (1-q)^{m-i} \quad 0 \leq q \leq 1;\ j = 0, 1, \ldots, m;\quad m = 1, 2, \ldots .$$

Then for a moderate sample size $n_i = n$ and $Z_i = z$, one can solve

$$1 - F_{z-1}(n,q_1) = \alpha/2$$

and (5)

$$F_z(n,q_2) = 1 - \alpha/2 \qquad \text{for } i=1,\ldots,r$$

for $p_i^*(n,z)$ and $p_i^{**}(n,z)$, respectively, and achieve a confidence coefficient of at least $1-\alpha$. We call the result a binomial interval.

As n_i increases, exact solution becomes difficult because of numerical error. Then one has the well known result

$$\lim_{n_i \to \infty} \text{pr}\left\{\frac{|\hat{p}_i - p_i|}{[p_i(1-p_i)/n_i]^{1/2}} \leq c(\alpha)\right\} = 1 - \alpha,$$

where

$$c(\alpha) = \left\{y: (2\pi)^{-1/2} \int_{-\infty}^{y} e^{-w^2/2} dw = 1-\alpha/2\right\},$$

and in principle one can solve the corresponding quadratic form

$$p_i^2[1+c^2(\alpha)/n_i] - p_i[2\hat{p}_i + c^2(\alpha)/n_i] + \hat{p}_i^2 = 0 \qquad (6)$$

for $p_i^*(n_i,Z_i) \leq p_i^{**}(n_i,Z_i)$. The resulting confidence interval has an error of approximation which decreases as n_i increases. However, the rate of convergence is nonuniform, being most rapid for $p_i = 1/2$ and least rapid for p_i close to zero and unity. This non-

uniform convergence limits the appeal of this confidence interval in practice.

A third approach uses Chebyshev's inequality so that $p_i^*(n_i,Z_i) \leq p_i^{**}(n_i,Z_i)$ are again solutions of (6), but with $1/\alpha^{1/2}$ replacing $c(\alpha)$. Although this confidence statement holds for every n_i, the interval width can be wide. A fourth approach based on the probability inequality (Okamoto 1955, Hoeffding 1963)

$$\Pr(Z_i - p_i \geq \varepsilon) \leq \left\{\left[p_i/(p_i+\varepsilon)\right]^{p_i+\varepsilon}\left[(1-p_i)/(1-p_i-\varepsilon)\right]^{1-p_i-\varepsilon}\right\}^{n_i}$$

$$0<\varepsilon\leq 1 - p_i$$

produces tighter intervals for small α. For $n_i \geq \ln(\alpha/2)/\ln \max(p_i,1-p_i)$, $(p_i^*(n_i,Z_i),\ p_i^{**}(n_i,Z_i))$ covers p_i with probability $> 1 - \alpha$ where $p_i^*(n_i,Z_i) \leq p_i^{**}(n_i,Z_i)$ are now the solutions to the equation

$$\hat{p}_i \ln(\theta/\hat{p}_i) + (1-\hat{p}_i) \ln[(1-\theta)/(1-\hat{p}_i)] = \frac{1}{n_i} \ln(\alpha/2). \quad (7)$$

See Fishman (1986). The ratio $\ln(\alpha/2)/\ln \max(\hat{p}_i,1-\hat{p}_i)$ provides an indication of whether or not n_i exceeds the required lower bound.

Although the resulting interval leads to a confidence interval of greater width than the binomial and normal intervals do, it is considerably easier to compute than the binomial interval is for moderate and large n_i and induces no error of approximation as the normal interval does. Therefore, we

recommend the computation of a binomial interval from (5) when possible and an interval based on (7) otherwise, provided that $n_i \geq \ln(\alpha/2)/\ln \max(\hat{p}_i, 1-\hat{p}_i)$.

Since $\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_r$ are independent, one has

$$\Pr\left\{\underset{\sim}{p} \varepsilon \prod_{i=1}^{r} [p_i^*(n_i, Z_i), p_i^{**}(n_i, Z_i)]\right\} \geq \beta$$

where $\beta = (1-\alpha)^r$. To achieve a confidence level β, one chooses $\alpha = 1 - \beta^{1/r}$ for each interval. Since the system is coherent, $\partial g(\underset{\sim}{p})/\partial p_i \geq 0$ for $i = 1, \ldots r$. Therefore

$$\Pr\left[g(\underset{\sim}{p}^*) \leq g(\underset{\sim}{p}) \leq g(\underset{\sim}{p}^{**})\right] \geq \beta \qquad (8)$$

where $\underset{\sim}{p}^* = (p_1^*(n_1, Z_1), \ldots, p_r^*(n_r, Z_r))$ and $\underset{\sim}{p}^{**} = (p_1^{**}(n_1, Z_1), \ldots, p_r^{**}(n_r, Z_r))$.

Since $p_i^*(n_i, Z_i) \leq \hat{p}_i \leq p_i^*(n_i, Z_i)$ for $i=1, \ldots, r$ with probability one and since $\partial g(\underset{\sim}{p})/\partial p_i \geq 0$ for $i=1, \ldots, r$, one has

$$g(\underset{\sim}{p}^*) \leq g(\hat{\underset{\sim}{p}}) \leq g(\underset{\sim}{p}^{**}) \qquad \text{with probability 1,}$$

a result which provides a convenient way of assessing the extent of sampling variation in $g(\hat{\underset{\sim}{p}})$. With $\underset{\sim}{p}^*$ and $\underset{\sim}{p}^{**}$ in hand for specified β, one can, for a specified system G, compute $g(\underset{\sim}{p}^*)$ and $g(\underset{\sim}{p}^{**})$ in two reliability evaluations and determine whether or not the interval width $g(\underset{\sim}{p}^{**}) - g(\underset{\sim}{p}^*)$ is sufficiently small for the purposes of reliability analysis. As Sections 3 and 4 show, there is good reason to believe that this interval grows

substantially as the size of the system G, constructed from components of types 1,...,r, grows.

3. Parallel and Series Systems

We use the s-t connectedness problem to illustrate the potential seriousness of errors in the estimate $g(\hat{\underset{\sim}{p}})$.

Theorem 1. Let G denote a network of k_1 arcs of type 1 in parallel with source node s and terminal node t so that

$$g(\underset{\sim}{p}) = 1 - (1-p_1)^{k_1} \tag{9}$$

gives the probability that s and t are connected. Let $Z_{11},\ldots,Z_{1n_1}$ denote 0-1 test data on n_1 components of this type, let $\hat{p}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Z_{ij}$, $\hat{\underset{\sim}{p}} = \hat{p}_1$ and $g(\hat{\underset{\sim}{p}}) = 1 - (1-\hat{p}_1)^{k_1}$. Also Chebyshev's inequality gives

$$\mathrm{pr}\{|\hat{p}_1 - p_1| < \beta[p_1(1-p_1)/n_1]^{1/2}\} > 1 - 1/\beta^2 \tag{10}$$

$\beta > 0.$

Based on (10), the minimal sample size required to achieve

$$\mathrm{pr}\{|g(\hat{\underset{\sim}{p}}) - g(\underset{\sim}{p})| < \varepsilon[1-g(\underset{\sim}{p})]\} > 1 - \frac{1}{\beta^2} \qquad \varepsilon>0 \tag{11a}$$

is

$$n_1^* > \beta^2 p_1/(1-p_1)[(1+\varepsilon)^{1/k_1}-1]^2 \qquad (11b)$$

and

$$\lim_{k_1 \to \infty} n_1^*/k_1^2 > \beta^2 p_1/[(1-p_1)[\ln(1+\varepsilon)]^2. \qquad (11c)$$

Proof. Substitution into (11a) gives

$$\text{pr}\{|g(\hat{\underset{\sim}{p}})-g(\underset{\sim}{p})|<\varepsilon[1-g(\underset{\sim}{p})]\} = \text{pr}\{[1-(1+\varepsilon)^{1/k_1}](1-p_1) \leq \hat{p}_1 - p_1 < [1-(1-\varepsilon)^{1/k_1}](1-p_1)\} > 1 - \frac{1}{\beta^2}.$$

Then Chebyshev's inequality (10) gives

$$n_1 > \beta^2 p_1/(1-p_1) \min\{[1-(1+\varepsilon)^{1/k_1}]^2, [1-(1-\varepsilon)^{1/k_1}]^2\}.$$

Since $(1+\varepsilon)^{1/k_1}-1 < 1-(1-\varepsilon)^{1/k_1}$, n_1^* in (11b) follows. Expression (11c) follows by applying L'Hopital's rule to $(1/k_1^2)/[(1+\varepsilon)^{1/k_1}-1]^2$.

Note that (11a) is an attempt to control the relative error on the system failure probability $1-g(\underset{\sim}{p})$. Expression (11c) immediately makes apparent that the sample size n_1^* needed to keep this relative error at ε grows quadratically with k_1, the number of arcs in parallel. Theorem 2 provides insight into the source of the potential error.

Theorem 2. For k_1 arcs of type 1 in parallel with $g(p_1) = 1 - (1-p_1)^{k_1}$ and $g(\hat{\underset{\sim}{p}}) = 1 - (1-\hat{p}_1)^{k_1}$,

$$\lim_{k_1 \to \infty} Eg(\hat{\underset{\sim}{p}}) = 1 - (1-p_1)^{n_1} \tag{12}$$

$$\lim_{n_1 \to \infty} n_1 E[g(\hat{\underset{\sim}{p}}) - g(\underset{\sim}{p})] = -k_1(k_1-1)p_1(1-p_1)^{k_1-1}/2 \tag{13}$$

and

$$\lim_{n_1 \to \infty} n_1 \text{ var } g(\hat{\underset{\sim}{p}}) = k_1^2 p_1(1-p_1)^{2k_1-1}. \tag{14}$$

Proof. Since

$$Eg(\hat{\underset{\sim}{p}}) = 1 - (1-p_1)^{n_1} - \sum_{j=0}^{n_1-1} (j/n_1)^{k_1} \binom{n_1}{j} p_1^j (1-p_1)^{n_1-j},$$

(12) follows immediately. Let $\Delta = \hat{p}_1 - p_1$ and observe that

$$g(\hat{\underset{\sim}{p}}) = 1 - (1-p_1-\Delta)^{k_1} = 1 - \sum_{j=0}^{k_1} \binom{k_1}{j} (1-p_1)^{k_1-j} \Delta^j (-1)^j$$

$$= 1 - (1-p_1-\Delta)^{k_1} + k_1(1-p_1)^{k_1-1}\Delta - k_1(k_1-1)(1-p_1)^{k_1-2}\Delta^2/2 + \ldots .$$

Since $E\Delta = 0$, $E\Delta^2 = p_1(1-p_1)/n_1$ and $E\Delta^m = O(1/n_1^{\lfloor (m+1)/2 \rfloor})$ $m \geq 3$,

$$E[g(\hat{\underset{\sim}{p}}) - g(\underset{\sim}{p})] = -k_1(k_1-1)p_1(1-p_1)^{k_1-1}/n_1 + O(1/n_1^2) \text{ as } n_1 \to \infty,$$

and (13) follows. An analogous development gives (14).

Theorem 2 raises several concerns. The quantity $g(\hat{\underset{\sim}{p}})$ understates the true reliability $g(\underset{\sim}{p})$. Moreover, the relative bias

$$\frac{E[g(\hat{\underset{\sim}{p}})-g(\underset{\sim}{p})]}{1 - g(\underset{\sim}{p})} \sim k_1(k_1-1)\, p_1/2(1-p_1)n_1$$

makes clear that the dominant term in the relative understatement increases quadratically in k_1. If the objective is to design a parallel system based on component of type 1 with a specified level of system reliability, then $g(\hat{\underset{\sim}{p}})$ encourages one to add more components in parallel than may truly be required.

Observe that the coefficient of variation

$$\gamma(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) = [\text{var } g(\hat{\underset{\sim}{p}})]^{1/2}/[1-g(\underset{\sim}{p})] \sim k_1[p_1/(1-p_1)n_1]^{1/2}$$

reveals linear growth in k_1. As a result, a network with Jk_1 components of type 1 in parallel would lead to a coefficient of variation J times larger than a network G_1 with just k_1 components in parallel.

An analogous development for $k_1 > 1$ components of type 1 in series gives a sample system reliability $g(\hat{\underset{\sim}{p}}) = \hat{p}_1^{k_1}$ that overstates the system reliability $g(\underset{\sim}{p}) = p_1^{k_1}$. Again, relative bias is proportional to k_1 and the coefficient of variation is proportional to k_1. More generally, consider a set of r subsystems in series where subsystem i is composed of $k_i > 1$ components of

type i in parallel i = 1,...,r. Here system reliability is $g(\underset{\sim}{p}) = \prod_{i=1}^{r} [1-(1-p_i)^{k_i}]$ and clearly the quantity $g(\hat{\underset{\sim}{p}}) = \prod_{i=1}^{r} [1-(1-\hat{p}_i^{k_i})]$ understates it. Conversely, a set of r subsystems in parallel where subsystem i has $k_i > 1$ components in series has reliability $g(\underset{\sim}{p}) = 1 - \prod_{i=1}^{r} (1-p_i^{k_i})$ and the quantity $g(\hat{\underset{\sim}{p}}) = 1 - \prod_{i=1}^{r} (1-\hat{p}_i^{k_i})$ overstates it.

4. More General Systems

Results for more general systems reveal how potential errors grow with the number of types of components r as well as with the number of components of each type.

Theorem 3. Consider a system composed of $k_1,\ldots,k_r$ components of types 1,...,r respectively. Then

$$Eg(\hat{\underset{\sim}{p}}) = g(\underset{\sim}{p}) + w(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) + R_1 \tag{15}$$

and

$$\text{var } g(\hat{\underset{\sim}{p}}) = v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) + R_2, \tag{16}$$

where

$$w(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) = \sum_{i=1}^{r} \{ \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x}) P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{p}) [\frac{k_i(k_i-1)p_i^2-2x_i(k_i-1)p_i+x_i(x_i-1)}{p_i^2(1-p_i)^2}] \} \frac{p_i(1-p_i)}{2n_i} \tag{17}$$

$$v(\underset{\sim}{k},\underset{\sim}{p},n) = \sum_{i=1}^{r} \{ \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x})\ P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{p})\ [\frac{x_i - k_i p_i}{p_i(1-p_i)}]\}^2\ \frac{p_i(1-p_i)}{n_i}\ , \tag{18}$$

$$R_1 = \sum_{i,j=1}^{r} O(1/n_i n_j) \qquad \text{as} \quad \min_{1\le i\le r} n_i \to \infty$$

and

$$R_2 = \sum_{i,j=1}^{r} O(1/n_i n_j), \qquad \text{as} \quad \min_{1\le i\le r} n_i \to \infty$$

Proof. Let $\Delta_i = \hat{p}_i - p_i$ and observe that

$$g(\underset{\sim}{\hat{p}}) = \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x})\ P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{\hat{p}})$$

$$= \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x}) \prod_{i=1}^{r} [\sum_{j=0}^{x_i} \sum_{m=0}^{k_i-x_i} \binom{x_i}{j}\binom{k_i-x_i}{m} p_i^{x_i-j}(1-p_i)^{k_i-x_i-m}(-1)^m \Delta_i^{j+m}]$$

$$= \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x}) \prod_{i=1}^{r} [p_i^{x_i}(1-p_i)^{k_i-x_i} \sum_{j=0}^{x_i} \sum_{m=0}^{k_i-x_i} \binom{x_i}{j}\binom{k_i-x_i}{m} \frac{(-1)^m \Delta_i^{j+m}}{p_i^j (1-p_i)^m}].$$

Expressions (15) and (16) following from substitution of (2), (3) and (4) for $E\Delta_i^{j+m}$ for $j,m=0,1,\ldots,k_i-x_i$ and the observation that that $E\Delta_i\Delta_{i'} = 0$ for $i\neq i'$.

In addition to the proportionality to $k_1^2,\ldots,k_r^2$, observe that the number of terms in $w(\underset{\sim}{k},\underset{\sim}{p},n)$ and $v(\underset{\sim}{k},\underset{\sim}{p},n)$ increases linearly with r, the number of types of components. This increase would become quadratic if the data vectors $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_r$ were positively correlated.

An alternative representation puts bias and variance into perspective with regard to the variation in $\{g(\underset{\sim}{p})\}$. Observe that

$$w(k,\underset{\sim}{p},n) = \sum_{i=1}^{r} [\partial^2 g(\underset{\sim}{p})/\partial p_i^2] p_i(1-p_i/n_i$$

and

$$v(k,\underset{\sim}{p},n) = \sum_{i=1}^{r} [\partial g(\underset{\sim}{p})/\partial p_i]^2 p_i(1-p_i)/n_i$$

where

$$\partial g(\underset{\sim}{p})/\partial p_i = \sum_{j \in \underline{E}_i} [g(1_{ij},\underset{\sim}{p}) - g(0_{ij},\underset{\sim}{p})], \tag{19}$$

$$\partial^2 g(\underset{\sim}{p})/\partial p_i^2 = \sum_{j \in \underline{E}_i} \sum_{\substack{k \in \underline{E}_i \\ k \neq j}} [g(1_{ij},1_{ik},\underset{\sim}{p}) - g(1_{ij},0_{ik},\underset{\sim}{p}) - g(0_{ij},1_{ik},\underset{\sim}{p}) \tag{20}$$

$$+ g(0_{ij},0_{ik},\underset{\sim}{p})]$$

and $g(a_{ij},\underset{\sim}{p})$ denotes reliability when $x_{ij} = a_{ij}$ and $g(a_{ij},a_{ik},\underset{\sim}{p})$ denotes reliability when $x_{ij} = a_{ij}$ and $x_{ik} = a_{ik}$ for $a_{ij},a_{ik} \in \{0,1\}$ and $j \neq k$.

5. Eliminating Bias

If, upon computation of the confidence interval in (8), one finds that the interval width $g(\underset{\sim}{p}^{**}) - g(\underset{\sim}{p}^{*})$ is within acceptable bounds, then the reliability point estimate $g(\hat{\underset{\sim}{p}})$ presumably meets the needs for analysis. When this is not so, one would like to find an improved estimate by reducing variance, reducing bias or reducing both. One approach increases the number of data points $n_1,\ldots,n_r$. A second approach, which we pursue here, looks for an

alternative method of using the current data more effectively.

Unfortunately, variance reduction is not possible. Since $\hat{\underset{\sim}{p}}$ is the maximum likelihood estimator of $\underset{\sim}{p}$, $g(\hat{\underset{\sim}{p}})$ is the maximum likelihood estimator of $g(\underset{\sim}{p})$ and $v(\underset{\sim}{k},\underset{\sim}{p},n)$ corresponds to the Cramer-Rao lower bound on variance for fixed $\underset{\sim}{k}$ and $\underset{\sim}{p}$ as $n_1,\ldots,n_r \to \infty$. That is, no alternative estimator of $g(\underset{\sim}{p})$ based on $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_r$ can achieve an asymptotic variance smaller than $v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})$ in (18).

The potential for bias removal is more promising. Recall that positive bias can lead to a more frequent failure pattern in practice than the computed reliability implies. A negative bias can lead to a costly enhancement of the system to mitigate the apparent, but not real, reliability deficit that the reliability computation suggests. This section describes a method of removing this statistical bias while preserving the asymptotic variance at its minimum $v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})$. The method uses a data resampling plan to produce an unbiased estimate of system reliability in time per trial that grows considerably more slowly than the time required to compute $g(\hat{\underset{\sim}{p}})$.

Recall that data vectors $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_r$ which were used to estimate $p_1,\ldots,p_r$ and assume that $n_i > k_i$ for $i = 1,\ldots,r$. Algorithm A describes a procedure that on each trial (step 2) randomly samples (without replacement) and assigns an element of $\underset{\sim}{Z}_i$ to each component of type i. Let $\underset{\sim}{X}$ denote the resulting arc state vector of zeros and ones. Given this assignment, the system either functions ($\phi(\underset{\sim}{X})=1$) or fails ($\phi(\underset{\sim}{X})=0$). Then $\hat{h}_K$ (step 3) is our refined measure of system reliability.

Algorithm A

Purpose: To compute an unbiased estimate $\hat{h}_K$ of system reliability $g(\underset{\sim}{p})$.

Input: Network $G = (\underline{V}, \underline{E})$, where $\underline{E} = \{e_{11},\ldots,e_{1k_1};\ldots; e_{r1},\ldots,e_{rk_r}\}$, sample data $\underset{\sim}{Z}_i = (Z_{i1},\ldots,Z_{in_i})$ $i = 1,\ldots,r$, and desired number of trials K.

Output: $\hat{h}_K$.

Nomenclature: $\underset{\sim}{X} = (X_{11},\ldots,X_{1k_1};\ldots,X_{r1},\ldots,X_{rk_r})$.

Method:

1. Set $S \leftarrow 0$.
2. On each of K trials:
   a. For $i = 1,\ldots,r$:

      $\underline{W}_i \leftarrow \{1,\ldots,n_i\}$.

      For $j = 1,\ldots,k_i$: sample e from $\underline{W}_i$; remove e from $\underline{W}_i$; set $X_{ij} \leftarrow Z_{ie}$.

   b. Determine $\phi(\underset{\sim}{X})$; set $S \leftarrow S + \phi(\underset{\sim}{X})$.
3. Compute reliability

   $\hat{h}_K \leftarrow S/K$.

End of procedure

Theorem 4. For $\hat{h}_K$ as computed in Algorithm A,

a. $E\hat{h}_K = g(\underset{\sim}{p})$

b. $\text{var } \hat{h}_K = g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]/K + [v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) + \sum_{i=1}^{r}\sum_{j=1}^{r} O(1/n_i n_j)](K-1)/K$

c. $\lim_{K\to\infty} \text{var } \hat{h}_K = v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) + \sum_{i,j=1}^{r} O(1/n_i n_j)$ as $\min_{1\le i\le r} n_i \to \infty$.

Proof of a. Observe that

$$\phi(\underset{\sim}{X}) = \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x}) \prod_{i=1}^{r} \prod_{j=1}^{k_i} X_{ij}^{x_{ij}}(1-X_{ij})^{1-x_{ij}}$$

where the X_{ij}'s are sampled in step 2a. Since sampling occurs without replacement on each trial

$$E\prod_{j=1}^{k_i}[X_{ij}^{x_{ij}}(1-X_{ij})^{1-x_{ij}}] = \prod_{j=1}^{k_i} E[X_{ij}^{x_{ij}}(1-X_{ij})^{1-x_{ij}}]$$

$$= p_i^{x_i}(1-p)^{k_i-x_i}.$$

Therefore,

$$E\phi(\underset{\sim}{X}) = g(\underset{\sim}{p})$$

and consequently $\hat{h}_K$ is unbiased. Also, since $\phi^2(\underset{\sim}{X}) = \phi(\underset{\sim}{X})$ $\text{var } \phi(\underset{\sim}{X}) = g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]$.

Proof of b. Let

$$\underset{\sim}{q} = (q_{11},\ldots,q_{1k_1};\ldots;q_{r1},\ldots,q_{rk_1})$$

and redefine the reliability function as

$$h(\underset{\sim}{q}) = \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x}) \prod_{i=1}^{r} \prod_{j=1}^{k_i} q_{ij}^{x_{ij}}(1-q_{ij})^{1-x_{ij}}.$$

Now write the Taylor expansion

$$h(\underset{\sim}{q}) = h(p_1,\ldots,p_1;\ldots;p_r,\ldots,p_r) + \sum_{i=1}^{r}\sum_{j=1}^{k_i} \partial h/\partial q_{ij}\,|_{q_{ij}=p_i}(q_{ij}-p_i) + R$$

where R denotes the remainder composed of higher-order cross-derivatives. Let $X_{ij}^{(y)}$ and $X_{ij}^{(z)}$ denote the assignments for arc e_{ij} on trials v and w respectively. Then for all $j_1,j_2 = 1,\ldots,k_1$

$$E[(X_{ij_1}^{(y)}-p_i)(X_{ij_2}^{(z)}-p_i)] = p_i(1-p_i) \qquad \text{if } y=z \text{ and } j_1=j_2$$

$$= p_i(1-p_i)/n_i \qquad \text{if } y\neq z$$

$$= 0 \qquad \text{otherwise.}$$

Let

$$\underset{\sim}{X}^{(y)} = (X_{11}^{(y)},\ldots,X_{1k_1}^{(y)};\ldots;X_{r1}^{(y)},\ldots,X_{rk_r}^{(y)})$$

and $$\Delta_{ij}^{(y)} = X_{ij}^{(y)} - q_{ij}.$$

Then

$$h(\underset{\sim}{X}^{(y)}) = \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} \phi(\underset{\sim}{x})P(\underset{\sim}{x},\underset{\sim}{k},\underset{\sim}{p}) \prod_{i=1}^{r} \prod_{j=1}^{k_i}[1+\Delta_{ij}^{(y)}(2x_{ij}-1)/p^{x_{ij}}(1-p_i)^{1-x_{ij}}]$$

Since

$$\partial g/\partial \underset{\sim}{p}_i = \sum_{j=1}^{k_i} \partial h/\partial q_{ij}\big|_{q_{ij}=p_i} ,$$

one has for $y \neq z$.

$$\operatorname{cov}[h(\underset{\sim}{X}^{(y)}), h(\underset{\sim}{X}^{(z)})] = \sum_{i=1}^{r} (\partial g/\partial p_i)^2 \, p_i(1-p_i)/n_i + \sum_{i,j=1}^{r} O(1/n_i n_j) \quad \text{as} \quad \min_{1\le i\le r} n_i \to \infty, \; y \neq z .$$

Since

$$v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) = \sum_{i=1}^{r} (\partial g/\partial p_i)^2 \, p_i(1-p_i)/n_i \quad \text{as} \quad \min_{1\le i\le r} n_i \to \infty,$$

the quantity

$$\hat{h}_K = \frac{1}{K} \sum_{y=1}^{K} h(\underset{\sim}{X}^{(y)})$$

has

$$\operatorname{var} \hat{h}_K = \operatorname{var} h(\underset{\sim}{X}^{(y)})/K + \sum_{y\neq z} \operatorname{cov}[h(\underset{\sim}{X}^{(y)}), h(\underset{\sim}{X}^{(z)})](K-1)/K$$

$$= g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]/K + [v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n}) + \sum_{i,j=1}^{r} O(1/n_i n_j)](K-1)/K \quad \text{as} \quad \min_{1\le i\le r} n_i \to \infty ,$$

which proves part b. Part c follows immediately.

The significance of Algorithm A is now apparent. The resampling scheme produces an unbiased estimate of $g(\underset{\sim}{p})$. As the number of trials K increases, the variance of $\hat{h}_K$ converges to a quantity whose dominant term is the Cramér-Rao lower

bound $v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})$. Moreover, in place of a direct calculation of the reliability $g(\hat{\underset{\sim}{p}})$, one computes $\phi(\underset{\sim}{X})$, in step 2b, K times. For s-t connectedness, a depth-first search algorithm computes $\phi(\underset{\sim}{X})$ in $O(\max(|\underline{V}|,|\underline{E}|)$ time. See Aho, Hopcroft and Ullman (1974). If G is a directed acyclic flow network with random binary capacities and

$$\phi(\underset{\sim}{x}) = \phi(\underset{\sim}{x},z) = 1 \quad \text{if maximal s-t flow} > z$$

$$= 0,$$

one can determine $\phi(\underset{\sim}{X})$ in $O(|\underline{V}|\log|\underline{V}|)$ time if G is planar. See Itai and Shiloach (1979). The fastest known algorithm for a nonplanar network takes $O(|\underline{V}|^3)$. See Malhotra, Kumar and Maheshwari (1978). These time complexities make clear that the cost of resampling per trial is generally incidental relative to the cost of performing the exact computation of $g(\underset{\sim}{p})$.

To bring $\text{var}\,\hat{h}_K$ to the neighborhood of $v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})$ one needs to make K sufficiently large to make $g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]$ small relatively to $v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})$. To assess when this occurs, one would need to observe $\text{var}\,\hat{h}_K$ as a function of K. This quantity is unknown; moreover, it is not possible with the sampling scheme of Algorithm A to compute a useful estimate of $\text{var}\,\hat{h}_K$.

One partial solution to the problem partitions the data. Let $m_1,\ldots,m_r$ denote integers such that $m_i > k_i$ for $i=1,\ldots,r$, and let $c=n_1/m_1=\ldots=n_r/m_r$ = integer. Then Algorithm B describes an alternative scheme that involves resampling K^* times from each of c partitions of the data $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_r$. Theorem 5 reveals the benefit of this method.

Algorithm B

Purpose: To compute an unbiased estimate $\bar{h}_K$ of system reliability $g(\underset{\sim}{p})$ and an unbiased estimate of var $\bar{h}_K$.

Input: Network $G = (\underline{V}, \underline{E})$ where $\underline{E} = \{e_{11},\ldots,e_{1k_1};\ldots; e_{r1},\ldots,e_{rk_r}\}$, sample data $\underset{\sim}{Z}_i = (Z_{i1},\ldots,Z_{in_i})$ $i=1,\ldots,r$, integers $c, m_1,\ldots,m_r$ and desired number of replications per partition K^*.

Output: $\bar{h}_K$ and $V(\bar{h}_K)$ as unbiased estimates of $g(\underset{\sim}{p})$ and var $\bar{h}_K$.

Nomenclature: $\underset{\sim}{X} = (X_{11},\ldots,X_{1k_1};\ldots,X_{r1},\ldots,X_{rk_r})$.

Method:

1. Set $K \leftarrow 0$; For $y=1,\ldots,c$: set $S_y \leftarrow 0$.
2. On each of K^* trials:

   For $y=1,\ldots,c$:

   For $i=1,\ldots,r$:

   $\underline{W}_i \leftarrow \{1,\ldots,m_i\}$; For $j=1,\ldots,k_i$: sample e from $\underline{W}_i$; remove e from $\underline{W}_i$; set $X_{ij} \leftarrow Z_{i,(y-1)m_i+e}$.

   Determine $\phi(\underset{\sim}{X})$.

   Set $S_y \leftarrow S_y + \phi(\underset{\sim}{X})$.
3. $K \leftarrow K + cK^*$.
4. Compute summary statistics

$$\bar{h}_K \leftarrow (S_1+\ldots+S_c)/K.$$

$$V(\bar{h}_K) \leftarrow \frac{1}{c(c-1)} \sum_{y=1}^{c} (cS_y/K-\bar{h}_K)^2.$$

End of procedure

Theorem 5. For the resampling scheme in Algorithm B,

a. $E\bar{h}_K = g(\underset{\sim}{p})$

b. $\text{var}\ \bar{h}_K = g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]/K + v(\underset{\sim}{k},\underset{\sim}{p},\underset{\sim}{n})\ (K-c)/K + c \sum_{i,j=1}^{r} O(1/n_i n_j)$

as $\min_{1\le i\le r} n_i \to \infty$

c. $EV(\bar{h}_K) = \text{var}\ \bar{h}_K$.

Proof. Within any partition y, the resampling scheme is identical with that of Algorithm A except that sampling occurs from $Z_{i,(y-1)m_i+1},\ldots,Z_{i,ym_i}$ for $i=1,\ldots,r$. Therefore, for each $y=1,\ldots,c$

$$E(S_y/K^*) = g(\underset{\sim}{p}),$$

establishing part a. Also

$$\text{var}(S_y/K^*) = g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]/K^* + \Big[\sum_{i=1}^{r} (\partial g/\partial p_i^2)\ p_i(1-p_i)/m_i + \sum_{i,j=1}^{r} O(1/m_i m_j)\Big]\ (K^*-1)/K^* \quad \text{as} \quad \min_{1\le i\le r} m_i \to \infty .$$

Since $S_1,\ldots,S_c$ are independent, one has

$$\text{var}\ \bar{h}_K(\underset{\sim}{p}) = \text{var}(S_y/K^*)/c.$$

Using this result, together with $m_i = n_i/c$ $i=1,\ldots,r$ and $K = cK^*$, gives part b. Part c follows by taking expectations.

The quantity $V(\bar{h}_K)$ provides a useful estimate of var $\bar{h}_K$ which one can use sequentially to estimate when the quantity $g(\underset{\sim}{p})[1-g(\underset{\sim}{p})]/K$ becomes relatively incidental to the variance. That is, the organization of Algorithm B enables one to iterate on step 2 to generate successive estimates $\bar{h}_{cK}*$, $\bar{h}_{2cK}*$,... and $V(\bar{h}_{cK}*)$, $V(\bar{h}_{2cK}*)$,... and observe the extent to which this variance measure stablizes as a function of K.

The one drawback of Algorithm B as compared to the Algorithm A arises from the increased relative importance of the higher order terms $\sum_{i,j=1}^{r} O(1/n_i n_j)$. These are scaled by c in Algorithm B. As the sample sizes $n_1,\ldots,n_r$ increase, these terms diminish in importance in each case, although they always remain c times larger in Algorithm B. In practice, as $m_1,\ldots,m_r$ increase c decreases, reducing the significance of the higher-order terms. However, a smaller c means a less statistically reliable estimate $V(\bar{h}_K)$ of var $\bar{h}_K$.

6. Conclusions

In general, the observations made in this paper are not encouraging about the statistical accuracy of a system reliability computation whose input consists of component reliability estimates. Although no alternative system reliability estimator produces a smaller asymptotic variance, the resampling schemes of Section 5 do provide a way of reducing bias. Based on the material presented here, a constructive approach to system reliability error assessment follows these steps:

1. Compute component reliability estimates $\hat{p}_1,\ldots,\hat{p}_r$.
2. Compute $100\times(1-\alpha)^{1/r}$ confidence intervals for each component reliability p_i for $i=1,\ldots,r$.
3. Compute a system reliability estimate using $\hat{p}_1,\ldots,\hat{p}_r$ as input.
4. Compute a $100\times(1-\alpha)$ confidence interval for system reliability using the confidence intervals for $p_1,\ldots,p_r$ in step 2.
5. If the interval width for system reliability is within acceptable bounds at coverage level $1-\alpha$, proceed with the study. Otherwise:
   a. One may improve the statistical accuracy of the point estimator by employing the resampling schemes in Section 6

      or

   b. One may wish to collect more test data to improve the component reliability estimates and thereby shorten the interval.

References

1. Aho, A.V., J.E. Hopcroft and J.D. Ullman (1974). The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Massachusetts.

2. Ball, M.O. and J.S. Provan (1983). Bounds on the reliability problem for shellable independence systems, SIAM J. Alg. and Disc. Meth., 3, 166-181.

3. Barlow, R.E. and F. Proschan (1981). Statistical Theory of Reliability and Life Testing Probability Models, To Begin With, Silver Spring, Maryland.

4. Fishman, G.S. (1986). Confidence intervals for mean and proportions in the bounded case, Technical Report No. UNC/ORSA/TR-86/19, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.

5. Gaver, D.P. and D.G. Hoel (1970). Comparison of certain small-sample Poisson probability estimates, Technometrics, 12, 835-850.

6. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc., 58, 13-29.

7. Itai, A. and Shiloach, Y. (1979). Maximum flow in planar networks, SIAM J. Comput., 8, 2, 135-150.

8. Malhotra, V.M., M.P. Kumar and S.N. Maheshwari (1978). An $O(|V|^3)$ algorithm for finding maximum flows in networks, Inf. Proc. Letters, 7, 277-278.

9. Okamoto, M. (1958). Some inequalities relating to the partial sum of binomial probabilities, Annals of the Inst. of Stat. Math., 10, 29-35.

10. Provan, J.S. (1986). The complexity of reliability computations in planar and acyclic graphs, SIAM J. Comp., 15, 694-702.

11. Valiant, L.G. (1979). The complexity of enumeration and reliability problems, SIAM J. Comp., 8, 410-421.

END

12-87

DTIC