9/4
12/29/78

LA-7556-MS

Informal Report

DR. 2072

MASTER

An Exploratory Comparison of Methods for Combining Failure-Rate Data from Different Data Sources

University of California

LOS ALAMOS SCIENTIFIC LABORATORY
Post Office Box 1663 Los Alamos, New Mexico 87545

LA-7556-MS
Informal Report

UC-79p (Base Technology)
Issued: November 1978

An Exploratory Comparison of Methods for Combining Failure-Rate Data from Different Data Sources

H. F. Martz
R. A. Waller

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

AN EXPLORATORY COMPARISON OF METHODS FOR COMBINING FAILURE-RATE DATA FROM DIFFERENT DATA SOURCES

by

H. F. Martz, Jr. and R. A. Waller

ABSTRACT

Thirteen methods are considered for use in pooling failure-rate data from different data sources. A Bayesian approach is taken in which two distinct sources of variation are assumed to be present; namely, prior variation between data sources and statistical error variation within each data source. An exploratory Monte Carlo simulation is used to compare the performance of the methods when used to construct both pooled point and 90% interval estimates of the failure-rate. The results indicated that those methods based on simple averaging techniques are satisfactory when only a small number of data sources are to be pooled. When there are fifteen or more data sets to be pooled, more sophisticated methods, which incorporate additional model structure, are superior. An example is given to illustrate the use of each of the proposed methods.

I. INTRODUCTION

Frequently, in analyzing reliability data, the reliability analyst is confronted with nonhomogeneous data that may be pooled or combined in some manner in order to produce better reliability estimates. Such data may be experimental data that has been collected under somewhat different experimental conditions; failure data obtained from different data sources; operational data derived from a population of plants that include plant environmental effects, etc. Breipohl[1] (1978) refers to the combined estimates as "group" or "joint" estimates, while an estimate based on each data source is referred to as an "individual" estimate. For example, a recent IEEE project* involved soliciting information

*IEEE Project 500, Subcommittee SC-5. Reliability, Power Engineering Society, Nuclear Power Engineering Committee, Institute of Electrical and Electronics Engineers, Inc.

from a group of over 200 experts and consultants on failure-rates for electrical, electronic, and sensing components used in nuclear power plants. The estimates supplied by the experts were pooled in order to obtain both single group point and interval estimates for the failure-rate of interest. IEEE Std 500-1977[2](1977) contains the resulting estimates.

Various methods can be postulated for pooling the individual estimates into a single group estimate and range in complexity from weighed averaging techniques to complex adjustment methods based on highly structured models. Several of these methods are presented in Sec. 3. Breipohl (1978) has shown that group predictions outperform individual predictions and that a weighed average is a satisfactory method of combining the individual estimates. It is shown in Sec. 4 that such is not always the case, depending upon the structure of the model from which the data are assumed to have been generated. Martz[3](1975) discusses the use of an empirical Bayes approach for pooling failure-rate data.

The methods for pooling the data depend upon the structure of the model underlying the data. For the purposes of this report, we assume a general Bayesian structure of the problem in which two distinct types of variation are assumed to be present. The first type of variation represents the inherent variability in the reliability parameter underlying the data among the population of data sources. This variation in the underlying parameter will be referred to as prior variation, since it is present regardless of whether or not estimates are supplied by each data source. The second type of variation represents variability in the estimate supplied by each data source from the true parameter value underlying the data source. This variation is due to the fact that, based on limited data, parameter values are never exactly known. It is referred to as statistical variation within each data source. Both components of variation are illustrated in Fig. 1, in which the reliability parameter of interest is a failure rate λ of some device. The $\hat{\lambda}$ values represent estimates of the corresponding λ values supplied by the data sources. The subscript indexes a particular data source among the N assumed data sources. Thus, deviations on the horizontal axis represent the prior variation in λ, while vertical departures from the line $\hat{\lambda} = \lambda$ represent the statistical variation in the corresponding $\hat{\lambda}$ value. It is observed that the total variation in the λ, values, represented by the vertical axis, includes both the prior and statistical variance components. We are concerned with the estimation of each of these components of variation.

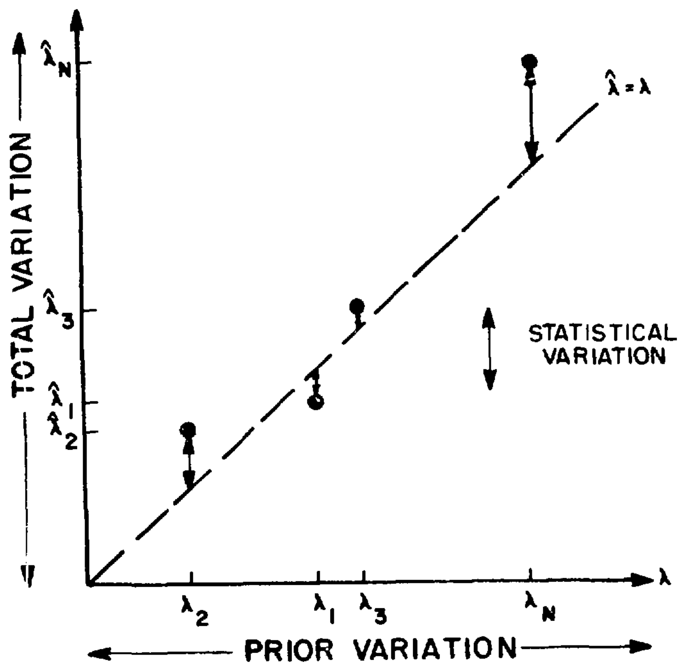

Fig. 1. The relationship between the prior, statistical, and total sources of variation.

One important fact should be noted. As the quantity of data available for a given data source increases, the statistical variation with respect to that source decreases. Further, as the quantity of data at each data source approaches infinity, the statistical variation at each data source approaches zero. Thus, the <u>individual</u> failure-rates λ_i would be estimated perfectly; however, any <u>group</u> estimate would still exhibit variability through the prior variation component. That is, an interval estimate based on large amounts of data from different sources need not approach zero width because the prior variation component is not diminished by increases in the respective sample sizes. In this case, the total variation would be equivalent to the prior variation. However, if no prior variation is present in the failure-rate population, then the width of the group interval estimate would tend toward zero as the quantity of data approaches infinity. This is precisely what happens when Bayesian interval estimates are constructed based on the posterior distribution of λ. Such Bayesian estimates are <u>not</u> group estimates and their width approaches zero, becoming more and more concentrated, as the amount of data upon which they are conditioned approaches infinity. This is an undesirable property because any group estimate should retain a prior variation component regardless of the quantity of data that is available to be pooled. Thus, the usual Bayesian interval estimates based on the posterior distribution are not suitable for direct use in pooling the individual estimates.

Let us now quantify some of the foregoing notions. For convenience, henceforth we will restrict our attention to the case in which a failure-rate is the desired reliability parameter of interest. We shall further assume that there are N individual data sources, or individual failure-rate estimates, that are to be pooled into a single group failure-rate estimate. Let λ_i, $i = 1,2,\ldots,N$, represent the true failure-rate corresponding to the i^{th} data source, and let $\hat{\lambda}_i$, $i = 1,2,\ldots,N$ represent the corresponding point estimate of λ_i supplied by the i^{th} data source. Based on the preceding discussion, we assume that λ_i is a value of a random variable λ according to some prior distribution $g(\lambda)$. It is also assumed that $\hat{\lambda}_i$, conditional on λ_i, is the value of a random variable according to some conditional distribution $f(\hat{\lambda}_i|\lambda_i)$. Thus, g expresses the prior variation in the parameter λ, while f accounts for the statistical variation in the estimate $\hat{\lambda}_i$. We shall further assume that, in addition to the best point estimate $\hat{\lambda}_i$ of λ_i supplied by the i^{th} data source, the i^{th} data source also supplies upper and lower bounds, denoted by $\hat{\lambda}_{i,U}$ and $\hat{\lambda}_{i,L}$, of λ_i. Further, in accordance with the IEEE Project 500 referenced earlier, let $r_i = 1,2,3,4,5$ denote the "expertise" self-rating associated with the i^{th} data source, where $r_i = 1$ corresponds to a low rating and $r_i = 5$ refers to a high rating. In general, r_i may be considered to be a "weight" assigned to the i^{th} data source.

In Sec. 2, several methods for pooling the individual failure-rate estimates into both group point and interval estimates are presented. An example is presented in Sec. 3. A Monte Carlo simulation is used to compare these methods and is discussed in Sec. 4. The results are also contained in Sec. 4. Finally, Sec. 5 presents the conclusions from this study.

II. POOLING FAILURE-RATE DATA

The authors participated in the IEEE Project 500 referenced in the preceding section. A part of the participation was to suggest a possible procedure for pooling individual estimates. The procedures suggested were then circulated for peer review. Response to the review provided numerous alternative procedures by various people. The first two methods presented below were the original methods suggested by the authors for use in pooling the individual estimates in the IEEE Project 500. Methods 3-6 are modifications to Method 1 and 2. The remaining seven methods were suggested by the respondents during the peer review of Methods 1 and 2.

Method 1: It is assumed that the conditional distribution $f(\hat{\lambda}_i|\lambda_i)$ has mean λ_i. In other words, the estimate $\hat{\lambda}_i$ supplied by the i^{th} data source is an unbiased estimate of the underlying failure-rate λ_i. It is further assumed that the standard deviation of $f(\hat{\lambda}_i|\lambda_i)$ is unknown and must be estimated. It is proposed to estimate this unknown standard deviation by R_i/d_i, where $R_i = \hat{\lambda}_{i,U} - \hat{\lambda}_{i,L}$ and where d_i is taken from E(W) in Table II of "Tables of Range and Studentized Range" by H. Leon Harter in the Annals of Mathematical Statistics, December 1960, pp. 1130-1131. The quantity d_i is chosen to reflect the degree of expertise associated with a data source. Therefore, it is associated with the expertise self-rating r_i. It may also be rationalized by using a sample size to correspond to a level of expertise. We have arbitrarily used the following assignment:

Let $r_i = 5$ correspond to $n = 25$ to give $d_i = 3.931$
$r_i = 4$ correspond to $n = 20$ to give $d_i = 3.735$
$r_i = 3$ correspond to $n = 15$ to give $d_i = 3.472$
$r_i = 2$ correspond to $n = 10$ to give $d_i = 3.078$
$r_i = 1$ correspond to $n = 5$ to give $d_i = 2.326$.

For example, if a data source rated himself at level four and supplied the upper and lower bounds 8×10^{-6} f/h and 4×10^{-7} f/h, respectively, we would take $7.6 \times 10^{-6}/3.735 = 2 \times 10^{-6}$ as our estimate of the standard deviation of the point estimate supplied by the data source.

Suppose we let θ and ψ denote the mean and standard deviation, respectively, of the prior distribution $g(\lambda)$. We desire to characterize the population of λ_i values by pooling the individual failure-rate estimates and to provide both a point and a probability interval estimate for λ. Both of these estimates will depend upon suitable estimates of θ and ψ.

It is proposed to estimate θ by use of a weighted average of the $\hat{\lambda}_i$ estimates

$$\hat{\theta} = \sum_{i=1}^{N} r_i \hat{\lambda}_i / \sum_{i=1}^{N} r_i \ . \tag{1}$$

This estimate will then be used as the group point estimate of the failure-rate random variable λ. In fact, $\hat{\theta}$ is the proposed point estimate in Methods 1-11 to be presented. Only in Methods 12 and 13 does the point estimate differ from Eq. (1).

Now let us consider an estimate of ψ, which is used in calculating the group probability interval estimate of λ. We use the well-known relationship between conditional and unconditional variance given by

$$\mathrm{Var}(\hat{\lambda}) = E[\mathrm{Var}(\hat{\lambda}|\lambda)] + \mathrm{Var}[E(\hat{\lambda}|\lambda)] \ . \tag{2}$$

Since it is assumed that $E(\hat{\lambda}|\lambda) = \lambda$, we find upon substitution that

$$\mathrm{Var}(\lambda) = \mathrm{Var}(\hat{\lambda}) - E[\mathrm{Var}(\hat{\lambda}|\lambda)] \ . \tag{3}$$

In regard to Fig. 1, this equation says in effect that the prior variation is equal to the total variation minus the average statistical variation.

Let us now consider estimates of each of the terms in Eq. (3). Now $\mathrm{Var}(\hat{\lambda})$ may be estimated by means of the weighted total sample variance of the $\hat{\lambda}_i$ values given by

$$V_1 = \sum_{i=1}^{N} (\hat{\lambda}_i - \theta)^2 \, r_i / \sum_{i=1}^{N} r_i \ , \tag{4}$$

and $E[\mathrm{Var}(\hat{\lambda}|\lambda)]$ may be estimated by means of the weighted average of the statistical variances given by

$$V_2 = \sum_{i=1}^{N} (R_i/d_i)^2 \, r_i / \sum_{i=1}^{N} r_i \ . \tag{5}$$

From Eq. (3), the estimate of the prior variance ψ^2 of λ is given as

$$\hat{\psi}^2 = V_1 - V_2 \ . \tag{6}$$

Now it may happen that $V_2 > V_1$, particularly when N is quite small. If this should occur, it is arbitrarily decided that $\hat{\psi}^2 = V_1$; that is, the prior variance will be estimated to be the total variation in the $\hat{\lambda}_i$ values. Another alternative in this case would be to set $\hat{\psi}^2$ equal to some arbitrarily small quantity ε. The choice to use V_1 rather than ε is motivated by similar results in empirical

Bayes decision theory and by practical observation. The empirical success of this choice is demonstrated in Sec. 4.

Chebychev's Inequality [Hogg and Craig[4] (1970), p. 55] is used to obtain the required probability interval (PI) estimate of λ as follows:

$$100(1 - \frac{1}{k^2})\% \text{ PI:} \quad \hat{\theta} - k\hat{\psi} \leq \lambda \leq \hat{\theta} + k\hat{\psi} , \tag{7}$$

where k is chosen to provide the desired containment probability. It is noted that this interval is such that the probability that it contains λ is <u>at least</u> $(1 - 1/k^2)$. It is known that such intervals are usually conservative in the sense that the actual containment probability is greater than $(1 - 1/k^2)$. For example, for an at least 90% probability interval on λ, we choose k = 3.16. It is also noted that this interval does not depend upon the form of $g(\lambda)$, provided that θ and ψ exist,which is assumed here. Thus, it is a prior distribution free method.

<u>Method 2</u>: The only difference between Method 2 and Method 1 is that in Method 2, a gamma prior distribution is assumed for λ and used to obtain the required probability interval estimate of λ rather than Chebychev's Inequality. Here we assume that λ has a gamma prior distribution given by

$$g(\lambda) = \frac{\lambda^{\alpha-1} e^{-\lambda/\beta}}{\beta^{\alpha} \Gamma(\alpha)} , \lambda, \alpha, \beta > 0 , \tag{8}$$

where α and β are the prior shape and scale parameters, respectively. Because $E(\lambda) = \theta = \alpha\beta$ and $Var(\lambda) = \psi^2 = \alpha\beta^2$, we can use the method of moments to estimate α and β by

$$\hat{\beta} = \hat{\psi}^2/\hat{\theta} \text{ and } \hat{\alpha} = \hat{\theta}^2/\hat{\psi}^2 . \tag{9}$$

The incomplete gamma function is then used to obtain the desired $100(1 - \gamma)\%$ probability interval for λ by finding the lower and upper limits λ_L and λ_U such that

$$P(\lambda < \lambda_L) = \int_0^{\lambda_L} g(\lambda)\, d\lambda = \gamma/2$$

and (10)

$$P(\lambda > \lambda_U) = 1 - \int_0^{\lambda_U} g(\lambda)\, d\lambda = \gamma/2 \ .$$

Thus, the required 100(1 - γ)% PI estimate for λ is given by

$$100(1 - \gamma)\%\ \text{PI:} \quad \lambda_L \le \lambda \le \lambda_U \ . \tag{11}$$

Method 3: This procedure differs from Method 1 in the manner used to estimate the unknown variance of $f(\hat{\lambda}_i|\lambda_i)$ for each data source. Rather than using the d_i factors based on the studentized range as in Method 1, Chebychev's Inequality is used to estimate the variance of the estimate supplied by each data source. It is assumed that the upper and lower limits $\hat{\lambda}_{i,U}$ and $\hat{\lambda}_{i,L}$, respectively, include at least 95% of the area under the conditional distribution $f(\hat{\lambda}_i|\lambda_i)$. It follows that in this case the standard deviation of $f(\hat{\lambda}_i|\lambda_i)$ may be estimated as $R_i/8.94$, where $R_i = \hat{\lambda}_{i,U} - \hat{\lambda}_{i,L}$. Thus, V_2 corresponding to Eq. (5) here becomes

$$V_2 = \sum_{i=1}^{N} (R_i/8.94)^2\, r_i / \sum_{i=1}^{N} r_i \ , \tag{12}$$

while V_1 is the same as in Eq. (4). The remaining steps in the method are the same as in Method 1 in which Chebychev's Inequality is used to obtain the required probability interval estimate. The details are given in Method 1.

Method 4: This method differs from Method 2 in the manner used to estimate the unknown variance of $f(\hat{\lambda}_i|\lambda_i)$ for each data source. As in Method 3, Chebychev's Inequality is used to obtain this estimate. The average variance estimate V_2 given in Eq. (12) is again used here. The remaining steps in this method are the same as in Method 2 in which a gamma prior distribution is assumed for

λ and used to provide the required probability interval estimate of λ. The details are given in Method 2.

Method 5: This method differs from Methods 1 and 3 in the manner used to estimate the conditional variance of $\hat{\lambda}_i|\lambda_i$, for each data source. In this method the upper and lower limits $\hat{\lambda}_{i,U}$ and $\hat{\lambda}_{i,L}$, respectively, for the i^{th} data source are assumed to be the extrema of a uniform distribution over this range. It follows that the standard deviation of $\hat{\lambda}_i$, conditional on λ_i, may be estimated as $R_i/\sqrt{12}$, where again $R_i = \hat{\lambda}_{i,U} - \hat{\lambda}_{i,L}$. Thus, V_2 corresponding to Eq. (5) now becomes

$$V_2 = \sum_{i=1}^{N} (R_i^2/12)\, r_i / \sum_{i=1}^{N} r_i \, , \tag{13}$$

while V_1 is the same as in Eq. (4). The remaining steps are the same as in Methods 1 and 3 in which Chebychev's Inequality is used to obtain the required probability interval estimate. Method 1 gives the details.

Method 6: This method is the same as Methods 2 and 4 except that a uniform distribution is used to estimate the conditional variance of $\hat{\lambda}_i|\lambda_i$ for each data source as in Method 5. The average variance estimate V_2 given in Eq. (13) is again used here. The remaining steps are the same as in Methods 2 and 4 in which a gamma prior distribution is assumed for $g(\lambda)$ and used to provide the required probability interval estimate of λ. Method 2 should be consulted for the details of this method.

Method 7: In this method, the failure-rate estimates $\hat{\lambda}_i$ are treated as though they are the "true" failure-rates λ_i and the statistical variation component is ignored. It follows that the prior variance ψ^2 is estimated by V_1 given in Eq. (4). The moment estimates of α and β given in Eq. (9) are used to fit a gamma prior distribution to the failure-rate population. The required probability interval is obtained by means of the procedure outlined in Method 2.

Method 8: In this conservative method, the standard deviation of $f(\hat{\lambda}_i|\lambda_i)$ is estimated by R_i/d_i as in Method 1 with the difference being in the choice of values for d_i. It is proposed to be conservative by using the following assignment:

Let $r_i = 5$ correspond to $n = 6$ to give $d_i = 2.534$

$r_i = 4$ correspond to $n = 5$ to give $d_i = 2.326$

$r_i = 3$ correspond to $n = 4$ to give $d_i = 2.059$

$r_i = 2$ correspond to $n = 3$ to give $d_i = 1.693$

$r_i = 1$ correspond to $n = 2$ to give $d_i = 1.128$

It is then proposed to estimate the prior variance ψ^2 as

$$\hat{\psi}^2 = V_1 + V_2 , \qquad (14)$$

where V_1 and V_2 are given by Eq. (4) and Eq. (5), respectively. The author of this method based this selection on the fact that ..."it is better to have the prior be as vague (large variance) as the data or lack thereof allows." Although Eq. (14) cannot be theoretically justified, it is in keeping with the conservative nature of this method. As in Method 1, Chebychev's Inequality is then used to provide the required probability interval estimate of λ.

Method 9: This method is the same as Method 8, except that a gamma prior distribution is fitted to the failure-rate population. The required probability interval is then obtained by means of the procedure outlined in Method 2.

Method 10: This method is similar to Method 1, except in the manner of computing the average statistical variance component $E[Var(\hat{\lambda}|\lambda)]$. It is argued in this method that even when all data sources provide the same range estimate, the contribution to Eq. (6) given by V_2 in Eq. (5) is not equal to the common value of all the data sources when the r_i values are different. This fact follows from the double accounting in the expression for V_2. It is proposed to rectify this shortcoming by redefining V_2 as

$$V_2 = \sum_{i=1}^{N} (R_i/d)^2 \, r_i / \sum_{i=1}^{N} r_i , \qquad (15)$$

where d is independent of each expert's rating and is a conversion factor from the range value to the variance, dependent upon the probability level associated with the range estimates. For example, if all data sources are estimating their respective ranges at the 90% level, then $R = 3.29\sigma$ and thus $d = 3.29$.

As in Methods 8 and 9, the prior variance would be estimated as the sum of V_1 and V_2, i.e., $\hat{\psi}^2 = V_1 + V_2$. A pooled range estimate $\hat{R}$ could be obtained by multiplying the pooled standard deviation $\sqrt{\hat{\psi}^2}$ by the conversion factor d.

Chebychev's Inequality is then used to provide the required probability interval estimate of λ as described in Method 1.

Method 11: This method is the same as Method 10, except that a gamma prior distribution is fitted to the failure-rate population. The procedure outlined in Method 2 is then used to obtain the required probability interval estimate.

Method 12: This method is the one finally adopted for pooling the failure-rate estimates in the IEEE Project 500. It is described in IEEE Std 500-1977 (1977), pp. 18-20. The group point estimate is given by the geometric average of the individual best estimates supplied by the data sources. That is, the prior mean θ is estimated as

$$\hat{\theta} = \left(\prod_{i=1}^{N} \hat{\lambda}_i \right)^{1/N} . \tag{16}$$

In a similar way, the upper and lower endpoints of the group interval estimate of λ are also given by the respective geometric averages of the upper and lower bounds of the data sources. That is,

$$\hat{\lambda}_{min} = \left(\prod_{i=1}^{N} \hat{\lambda}_{i,L} \right)^{1/N}, \quad \hat{\lambda}_{max} = \left(\prod_{i=1}^{N} \hat{\lambda}_{i,U} \right)^{1/N} . \tag{17}$$

Method 13: This method is the same as Method 12, except in the way in which the group interval estimate of λ is computed. Instead of using geometric averages, it is proposed to use

$$\hat{\lambda}_{min} = \min_{i=1,2,\ldots,N} (\hat{\lambda}_{i,L}), \quad \hat{\lambda}_{max} = \max_{i=1,2,\ldots,N} (\hat{\lambda}_{i,U}), \tag{18}$$

as the lower and upper endpoints, respectively.

Table I gives a summary of the thirteen methods just discussed, where the numbers in parenthesis refer to equation numbers in the text.

III. EXAMPLE

In order to illustrate each of the methods presented in the preceding section, consider the following example. Table II presents a hypothetical set

TABLE I

A SUMMARY OF THE THIRTEEN SUGGESTED METHODS FOR COMBINING FAILURE-RATE DATA

Method	$\hat{\theta}$	V_1	V_2	$\hat{\psi}^2$	Interval Estimate	Comments
1	A[a]	(4)	(5)	(4)-(5)	Chebychev's Inequality	
2	A	(4)	(5)	(4)-(5)	Gamma Prior	
3	A	(4)	(12)	(4)-(12)	Chebychev's Inequality	
4	A	(4)	(12)	(4)-(12)	Gamma Prior	
5	A	(4)	(13)	(4)-(13)	Chebychev's Inequality	
6	A	(4)	(13)	(4)-(13)	Gamma Prior	
7	A	(4)	-	(4)	Gamma Prior	
8	A	(4)	(5)	(4)+(5)	Chebychev's Inequality	Different Rating Scheme
9	A	(4)	(5)	(4)+(5)	Gamma Prior	Different Rating Scheme
10	A	(4)	(15)	(4)+(15)	Chebychev's Inequality	
11	A	(4)	(15)	(4)+(15)	Gamma Prior	
12	G[a]	-	-	-	(17)	
13	G[a]	-	-	-	(18)	

[a]A-Weighted Arithmetric Average; G-Geometric Average.

TABLE II

HYPOTHETICAL EXAMPLE DATA SET

Data Source (i)	$\hat{\lambda}_i \times 10^6$	$\hat{\lambda}_{i,L} \times 10^6$	$\hat{\lambda}_{i,U} \times 10^6$	r_i
1	3	0	8	5
2	8	2	11	4
3	1	0	3	5
4	9	2	14	4
5	11	3	18	3
6	4	3	10	4
7	7	1	10	4
8	4	0	7	4
9	7	4	9	5
10	11	4	20	4

of data used to illustrate the calculations involved in each of the thirteen methods. The point estimate of the prior mean θ for Methods 1-11 is the pooled estimate given in Eq. (1). Here,

$$\hat{\theta} = \frac{5(3) + 4(8) + \ldots + 4(11)}{42} = 6.19 \times 10^{-6} \text{ f/h},$$

where, for convenience, we have included the "units" of 10^{-6} f/h only on the RHS of the equation. This convention will be followed throughout this section.

Also, since V_1 given in Eq. (4) is used in many of the methods, it is computed to be

$$V_1 = \frac{5(3 - 6.19)^2 + 4(8 - 6.19)^2 + \ldots + 4(11 - 6.19)^2}{42} = 10.39 \times 10^{-12}\ f^2/h^2 \ .$$

Now let us proceed with the calculations for each of the methods.

Method 1: Table III presents the required information for computing V_2 according to Eq. (5). Thus,

$$V_2 = \frac{5(2.035)^2 + 4(2.410)^2 + \ldots + 4(4.284)^2}{42} = 6.59 \times 10^{-12}\ f^2/h^2 \ .$$

Hence, the estimate of the prior variance ψ^2 becomes

$$\hat{\psi}^2 = 10.39 - 6.59 = 3.80 \times 10^{-12}\ f^2/h^2 \ .$$

For an at least 90% probability interval estimate on λ, we choose k in Eq. (7) such that $1 - 1/k^2 = 0.90$, that is, $k = 3.16$. Hence, the 90% pooled interval estimate of λ becomes

$$6.19 - 3.16(1.95) \le \lambda \le 6.19 + 3.16(1.95)$$

or

$$0.03 \times 10^{-6}\ f/h \le \lambda \le 12.35 \times 10^{-6}\ f/h \ .$$

Method 2: Assuming now that the prior distribution is a member of the gamma family of distributions given in Eq. (8), the parameters α and β are estimated as

$$\hat{\beta} = 3.80/6.19 = 0.61 \times 10^{-6}\ f/h$$

and

$$\hat{\alpha} = (6.19)^2/3.80 = 10.08 \ .$$

TABLE III

INFORMATION REQUIRED FOR COMPUTING V_2 IN METHOD 1

Data Source (i)	r_i	d_i	$R_i \times 10^6$	$R_i/d_i \times 10^6$
1	5	3.931	8	2.035
2	4	3.735	9	2.410
3	5	3.931	3	0.763
4	4	3.735	12	3.213
5	3	3.472	15	4.320
6	4	3.735	7	1.874
7	4	3.735	9	2.410
8	4	3.735	7	1.874
9	5	3.931	5	1.272
10	4	3.735	16	4.284

Thus, using an incomplete gamma function computer subroutine, the 90% interval estimate of λ is obtained by solving Eq. (10) with $\gamma = 0.10$ and is found to be

$$3.35 \times 10^{-6} \text{ f/h} \leq \lambda \leq 9.64 \times 10^{-6} \text{ f/h} .$$

Method 3: According to Eq. (12), V_2 here becomes

$$V_2 = \frac{5(8/8.94)^2 + 4(9/8.94)^2 + \ldots + 4(16/8.94)^2}{42} = 1.23 \times 10^{-12} \text{ f}^2/\text{h}^2 .$$

Hence,

$$\hat{\psi}^2 = 10.39 - 1.23 = 9.16 \times 10^{-12} \text{ f}^2/\text{h}^2 .$$

Based on Chebychev's Inequality, the at least 90% probability interval estimate on λ becomes

$$6.19 - 3.16(3.03) \leq \lambda \leq 6.19 + 3.16(3.03)$$

or

$$0 \leq \lambda \leq 15.77 \times 10^{-6} \text{ f/h} .$$

Method 4: Now, the gamma prior distribution has parameters that are estimated as

$$\hat{\beta} = 9.16/6.19 = 1.48 \times 10^{-6} \text{ f/h}$$

and

$$\hat{\alpha} = (6.19)^2/9.16 = 4.18 \,.$$

Thus, using an incomplete gamma function computer subroutine, the 90% interval estimate becomes

$$2.18 \times 10^{-6} \text{ f/h} \le \lambda \le 11.85 \times 10^{-6} \text{ f/h} \,.$$

Method 5: Here V_2 is computed according to Eq. (13) and becomes

$$V_2 = \frac{5(8^2/12) + 4(9^2/12) + \ldots + 4(16^2/12)}{42} = 8.55 \times 10^{-12} \text{ f}^2/\text{h}^2 \,.$$

Thus,

$$\hat{\psi}^2 = 10.39 - 8.55 = 1.84 \times 10^{-12} \text{ f}^2/\text{h}^2 \,.$$

Again using Chebychev's Inequality, the at least 90% probability interval estimate on λ becomes

$$6.19 - 3.16(1.36) \le \lambda \le 6.19 + 3.16(1.36)$$

or

$$1.89 \times 10^{-6} \text{ f/h} \le \lambda \le 10.49 \times 10^{-6} \text{ f/h} \,.$$

Method 6: The gamma prior distribution has parameters that are estimated to be

$$\hat{\beta} = 1.84/6.19 = 0.30 \times 10^{-6} \text{ f/h}$$

and

$$\hat{\alpha} = (6.19)^2/1.84 = 20.82 \ .$$

Thus, the 90% interval estimate on λ becomes

$$4.18 \times 10^{-6} \text{ f/h} \leq \lambda \leq 8.65 \times 10^{-6} \text{ f/h} \ .$$

Method 7: Here,

$$\hat{\psi}^2 \equiv V_1 = 10.39 \times 10^{-12} \text{ f}^2/\text{h}^2 \ .$$

Thus, the parameters of the gamma prior distribution are estimated to be

$$\hat{\beta} = 10.39/6.19 = 16.8 \times 10^{-6} \text{ f/h}$$

and

$$\hat{\alpha} = (6.19)^2/10.39 = 3.69 \ .$$

The 90% interval estimate on λ becomes

$$2.00 \times 10^{-6} \text{ f/h} \leq \lambda \leq 12.28 \times 10^{-6} \text{ f/h} \ .$$

Method 8: For this method, $V_1 = 10.39 \times 10^{-12}$ f^2/h^2 and

$$V_2 = \frac{5(8/2.534)^2 + 4(9/2.326)^2 + \ldots + 4(16/2.326)^2}{42} = 17.23 \times 10^{-12} \text{ f}^2/\text{h}^2 \ .$$

The estimate of the prior variance ψ^2 becomes

$$\hat{\psi}^2 = 10.39 + 17.23 = 27.62 \times 10^{-12}\ f^2/h^2 .$$

Using Chebychev's Inequality, the <u>at least</u> 90% probability interval estimate on λ becomes

$$6.19 - 3.16(5.26) \leq \lambda \leq 6.19 + 3.16(5.26)$$

or

$$0 \leq \lambda \leq 22.81 \times 10^{-6}\ f/h .$$

<u>Method 9</u>: The parameters of the gamma prior distribution are estimated to be

$$\hat{\beta} = 22.62/6.19 = 4.46 \times 10^{-6}\ f/h$$

and

$$\hat{\alpha} = (6.19)^2/27.62 = 1.39 .$$

Thus, the 90% interval estimate on λ becomes

$$0.64 \times 10^{-6}\ f/h \leq \lambda \leq 16.57 \times 10^{-6}\ f/h .$$

<u>Method 10</u>: Now,

$$V_2 = \frac{5(8/3.29)^2 + 4(9/3.29)^2 + \ldots + 4(16/3.29)^2}{42} = 8.37 \times 10^{-12}\ f^2/h^2$$

from which

$$\hat{\psi}^2 = 10.39 + 8.37 = 18.76 \times 10^{-12}\ f^2/h^2 .$$

Using Chebychev's Inequality, an <u>at least</u> 90% probability interval estimate on λ is given by

$$6.19 - 3.16(4.33) \leq \lambda \leq 6.19 + 3.16(4.33)$$

or

$$0 \leq \lambda \leq 19.88 \times 10^{-6} \text{ f/h} .$$

Method 11: The parameters of the gamma prior distribution are estimated to be

$$\hat{\beta} = 18.76/6.19 \approx 3.03 \times 10^{-6} \text{ f/h}$$

and

$$\hat{\alpha} = (6.19)^2/18.76 = 2.04 ,$$

from which the required 90% interval estimate on λ is given by

$$1.13 \times 10^{-6} \text{ f/h} \leq \lambda \leq 14.57 \times 10^{-6} \text{ f/h} .$$

Method 12: The estimate of the prior mean θ is calculated from Eq. (16) as

$$\hat{\theta} = \left\{(3 \times 10^{-6})(8 \times 10^{-6})\ldots(11 \times 10^{-6})\right\}^{1/10} = 5.38 \times 10^{-6} \text{ f/h} .$$

Similarly, the limits of the interval estimate are computed according to Eq. (17) as

$$\hat{\lambda}_{min} = \left\{(0 \times 10^{-6})(2 \times 10^{-6})\ldots(4 \times 10^{-6})\right\}^{1/10} = 0$$

and

$$\hat{\lambda}_{max} = (8 \times 10^{-6})(11 \times 10^{-6})\ldots(20 \times 10^{-6})^{1/10} = 9.83 \times 10^{-6} \text{ f/h} .$$

Thus, the interval estimate of λ is

$$0 \leq \lambda \leq 9.83 \times 10^{-6} \text{ f/h} .$$

Method 13: The interval estimate is easily found according to Eq. (18) to be

$$0 \leq \lambda \leq 20.00 \times 10^{-6} \text{ f/h} .$$

Based on the data in Table II, Table IV gives the point and interval estimates for all thirteen methods. It is observed that Method 8 yields the most conservative, while Method 6 yields the least conservative, interval estimate.

IV. MONTE CARLO SIMULATION

A Monte Carlo FORTRAN simulation was conducted in order to assess and compare the performance of all thirteen methods for pooling individual failure-rate estimates into a single group point and interval estimate. The simulation was performed on a CDC 6600 Computer at the Los Alamos Scientific Laboratory. Sample sizes N of 2, 5, 10, 15, and 25 were used. Each replication in the simulation proceeded in the following manner. For the i^{th} data source, i=1,2,...,N, a "true" failure-rate λ_i was randomly drawn from a gamma prior distribution given by Eq. (8) with shape parameter α and scale parameter β. Three prior distributions were used; namely, $\alpha = 0.25$, $\beta = 24.0 \times 10^{-6}$h; $\alpha = 1.0$, $\beta = 6.0 \times 10^{-6}$h and

TABLE IV

A SUMMARY OF THE CORRESPONDING POINT AND INTERVAL ESTIMATES

Method	Point Estimate	Interval Estimate x 10^6
1	6.19×10^{-6} f/h	(0.03, 12.35)
2		(3.35, 9.64)
3		(0, 15.77)
4		(2.18, 11.85)
5		(1.89, 10.49)
6		(4.18, 8.65)
7		(2.00, 12.28)
8		(0, 22.81)
9		(0.64, 16.57)
10		(0, 19.88)
11		(1.13, 14.57)
12	5.38×10^{-6} f/h	(0, 9.83)
13		(0, 20.00)

$\alpha = 10.0$, $\beta = 0.6 \times 10^{-6}$ h. All three distributions have a prior mean of 6.0×10^{-6} f/h, although they have different "shapes." The first gamma prior is L-shaped with a standard deviation of 12.0×10^{-6} f/h; the second is an exponential distribution with a standard deviation of 6.0×10^{-6} f/h; and the third is a positively skewed unimodal model with a standard deviation of 1.9×10^{-6} f/h. Thus, the first distribution is significantly more diffuse than the third.

The estimate supplied by the i^{th} data source was simulated in the following way. Another gamma distribution was used in which the shape and scale parameters are denoted by α_i* and β_i*, respectively. Now α_i* was randomly sampled from a uniform distribution with range $[\alpha' - \varepsilon, \alpha' + \varepsilon]$. In this way, the shape parameter for the estimate supplied by the i^{th} data source could be made to either conform closely or not conform closely with the shape parameter value α of the prior distribution by taking $\alpha' \equiv \alpha$ and ε small or large, respectively. In all *simulation cases,* $\alpha' \equiv \alpha$ *and* ε *was set equal to either* 0.05 (*for* $\alpha = 0.25$ *or* 10.0), 0.125 (for $\alpha = 0.25$), or 4.0 (for $\alpha = 10.0$). In one case, $\varepsilon = 0.20$ was used. For example, the case in which $\alpha' = 10.0$ and $\varepsilon = 0.05$ would simulate the situation in which all data sources have information which is quite consistent with the underlying prior distribution. On the other hand, the case where $\alpha' = 10.0$ and $\varepsilon = 4.0$ simulates less consistent information about the prior distribution since the shape parameter underlying the estimate for each data source is expected to vary more widely than in the preceding case. Once the shape parameter value α_i* has been randomly selected, the corresponding scale parameter value β_i* was computed as

$$\beta_i^* = \lambda_i / \alpha_i^* \ .$$

Since it is assumed that each source reports an unbiased estimate of λ_i, by computing β_i* in this way it is insured that $E(\hat{\lambda}_i | \lambda_i) = \alpha_i^* \, \beta_i^* = \lambda_i$. Thus, each estimate $\hat{\lambda}_i$ is unbiased. The point estimate $\hat{\lambda}_i$ supplied by the i^{th} data source was then randomly drawn from a gamma distribution with parameters α_i* and β_i*.

The individual interval estimate supplied by the i^{th} data source was simulated as follows: Lower and upper $\gamma/2$ and $1 - \gamma/2$ percentiles of the gamma distribution with parameters α_i* and β_i* were computed and taken as the lower and upper bounds, $\hat{\lambda}_{i,L}$ and $\hat{\lambda}_{i,U}$, respectively. Three values of γ were used,

γ = 0.01, 0.10, and 0.20, in order to determine the effect of the individual interval width and the probability of coverage on the 90% group interval estimate. All data sources were considered to be reporting intervals with the same value of γ in each simulation run.

The results of the IEEE Project 500 revealed that most data sources rated themself either a four or five, with five being the most prevalent. In conformance to this observation, ratings were randomly assigned to each data source according to the following distribution: $P(r_i = 5) = 0.7$, $P(r_i = 4) = 0.2$, $P(r_i = 3) = 0.08$, $P(r_i = 2) = 0.01$, and $P(r_i = 1) = 0.01$.

After the data for the N data sources were simulated as described above, the thirteen pooling methods presented in Sec. 2 were used to compute both a single group point and a 90% interval estimate. The estimates were then compared to corresponding "true" values from the underlying prior distribution. One hundred independent repetitions of the above simulation experiment were made for each parameter combination. Several performance measures were defined; namely, the average point estimate, the average squared error of each point estimate and the "true" prior mean, and the standard error of each of these averages. The average and average squared error of the area under the true prior distribution within each group interval estimate was also computed. The closer this average coverage probability to 90%, the better the performance of the estimator. Also, the smaller the average squared error, the better the estimator's performance. The standard error of each of these averages was also computed. In many cases, the standard errors were less than 10% of the averages. In cases where conflicts arose regarding which method was superior, those cases were rerun to 1000 replications in order to further decrease the magnitude of the standard errors.

V. PERFORMANCE COMPARISONS

Twenty-seven simulation runs were made by varying certain parameters in each run. The basic experimental design for the simulation study consisted of the vertices of two cubes, since four parameters were varied in the study. Based on the preliminary results of these 16 runs, additional runs were made at the midpoint of the faces of one cube, at the center of the cube, and a few other selected positions as well. Table V gives the parameter combinations for the 27 runs that were made. These 27 combinations are graphically represented in Fig. 2 and indexed by case number. It is to be noted in Fig. 2 that

TABLE V

CHOICE OF PARAMETERS USED IN THE SIMULATION STUDY

Case Number	N	α	β	α'	ϵ	γ
1	5	0.25	24.0×10^{-6}	0.25	0.05	0.01
2	5	0.25	24.0×10^{-6}	0.25	0.05	0.10
3	15	0.25	24.0×10^{-6}	0.25	0.05	0.01
4	15	0.25	24.0×10^{-6}	0.25	0.05	0.10
5	5	0.25	24.0×10^{-6}	0.25	0.125	0.01
6	5	0.25	24.0×10^{-6}	0.25	0.125	0.10
7	15	0.25	24.0×10^{-6}	0.25	0.125	0.01
8	15	0.25	24.0×10^{-6}	0.25	0.125	0.10
9	5	10.0	0.6×10^{-6}	10.0	0.05	0.01
10	5	10.0	0.6×10^{-6}	10.0	0.05	0.10
11	15	10.0	0.6×10^{-6}	10.0	0.05	0.01
12	15	10.0	0.6×10^{-6}	10.0	0.05	0.10
13	5	10.0	0.6×10^{-6}	10.0	4.0	0.01
14	5	10.0	0.6×10^{-6}	10.0	4.0	0.10
15	15	10.0	0.6×10^{-6}	10.0	4.0	0.01
16	15	10.0	0.6×10^{-6}	10.0	4.0	0.10
17	15	1.0	6.0×10^{-6}	1.0	0.05	0.05
18	10	10.0	0.6×10^{-6}	10.0	0.05	0.05
19	5	1.0	6.0×10^{-6}	1.0	0.05	0.05
20	10	1.0	6.0×10^{-6}	1.0	0.05	0.10
21	10	0.25	24.0×10^{-6}	0.25	0.05	0.05
22	10	1.0	6.0×10^{-6}	1.0	0.05	0.01
23	10	1.0	6.0×10^{-6}	1.0	0.05	0.05
24	25	0.25	24.0×10^{-6}	0.25	0.05	0.01
25	15	0.25	24.0×10^{-6}	0.25	0.05	0.20
26	5	0.25	24.0×10^{-6}	0.25	0.05	0.20
27	2	0.25	24.0×10^{-6}	0.25	0.05	0.01

case numbers to the right of a point refer to the "faces," and "center," of the design. This convention will also be used in the remaining figures in this section. Table VI gives the values associated with the low, medium, and high levels of each of the parameters. Other values are indicated in Fig. 2.

TABLE VI

PARAMETER LEVELS USED IN THE SIMULATION STUDY

Parameter	Low Level	Medium Level	High Level
N	5	10	15
α	0.25	1.0	10.0
α'	0.25	1.0	10.0
ϵ	0.05	0.125	4.0
γ	0.01	0.05	0.10

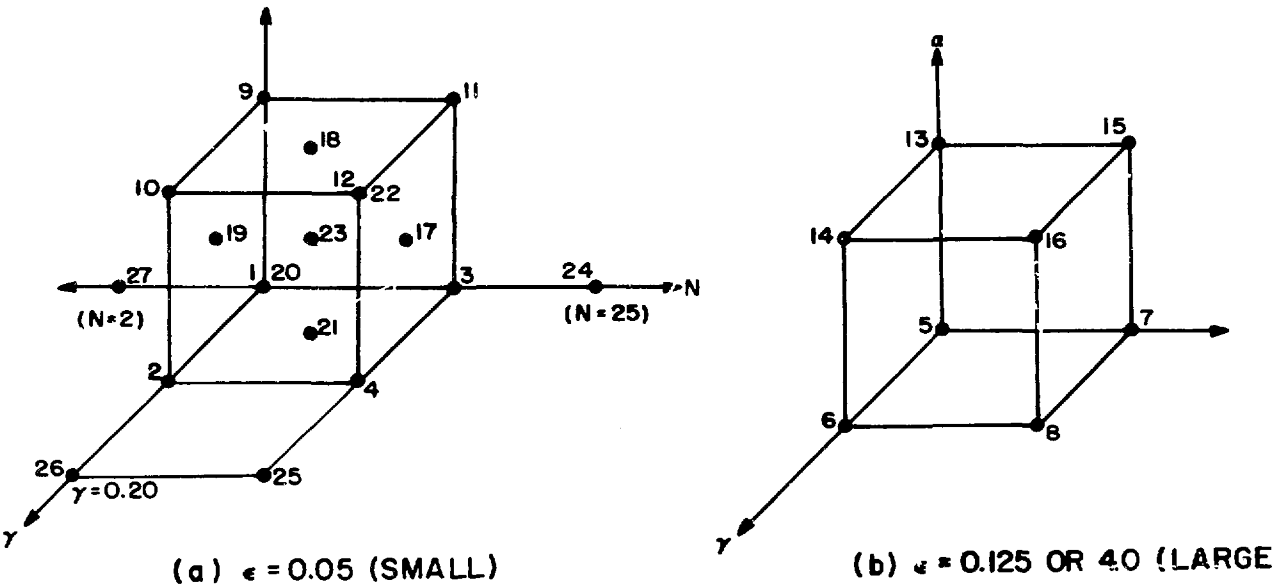

Fig. 2. Graphical representation of the twenty-seven cases in the simulation study.

First, let us consider the group interval estimate for the failure-rate population. Figure 3 gives the average (over the 100 replications) coverage probability (in percent) within the 90% theoretical interval for pooling method number 1 for all 27 simulated cases. The closer to 90%, the better a method performs when used to construct a 90% group interval estimate from the individual estimates. It is immediately observed in Fig. 3 that ε has practically no effect on the average coverage within the true 90% interval for Method 1. Recall that ε controls, in some sense, the degree of conformance of the shape of the gamma distribution of the data sources to the prior distribution. Thus, the variation in the shape of the gamma data source distributions has little if any effect on the average performance of the method, at least within the range of ε considered here. This was also found to be the case for the remaining methods and for all other performance measures as well. Henceforth, ε will be eliminated as a parameter in presenting the results of the simulation. Figures 4-15 give the average coverage probability within the 90% theoretical interval for pooling Methods 2-13. Since ε has little if any effect, the results presented are for those cases in Fig. 2(a) only. The conclusions based upon Figs. 3-15 are presented in the next section. Table VII presents the typical output of a

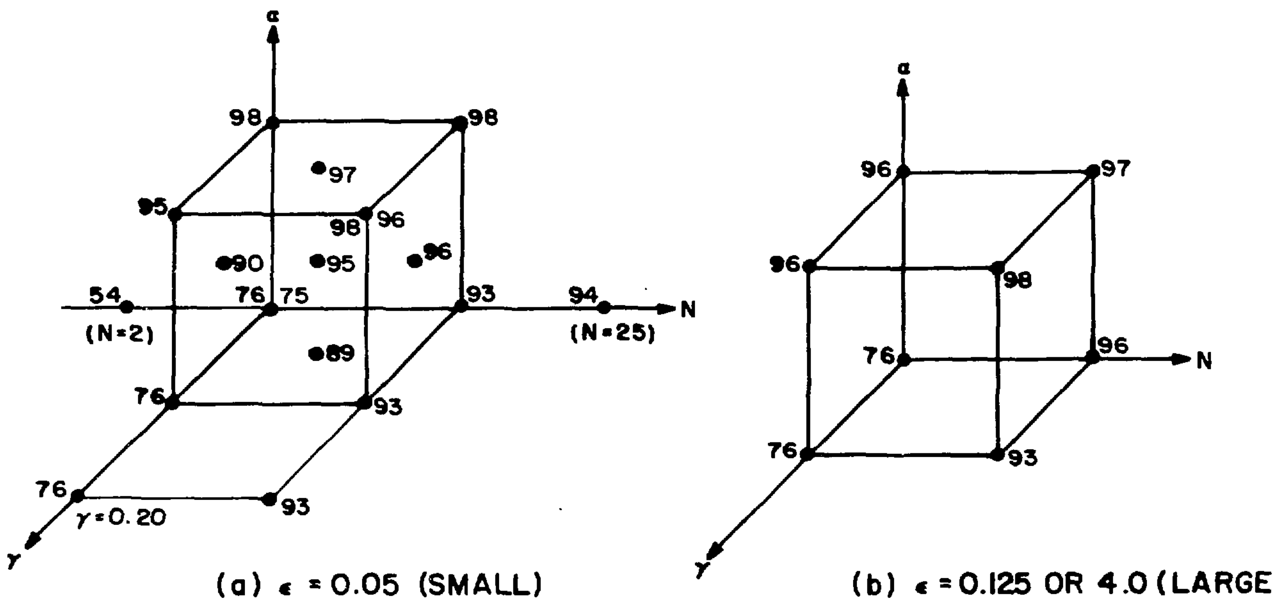

Fig. 3. Average coverage within the 90% theoretical interval for Method 1.

simulation run. The output is for case number 1, for the average, average squared error, and corresponding standard errors of the coverage probability within the 90% theoretical interval. The column headed AVERAGE in 27 tables such as these were used in constructing Figs. 3-15. Now, the method with the average closest to 90% was identified in all 27 runs and the results are presented in Fig. 16 for the 19 runs in Fig. 2(a). In a similar way, the method with the minimum average squared error over the 100 replications in each of the same 19 runs is presented in Fig. 17. For example, in Table VII, Method 13 has the minimum average squared error and this is indicated in Fig. 17 at the origin (case number 1). For N = 2, Methods 8, 10, and 13 were essentially tied for the minimum and all three methods are indicated.

Now let us consider the group point estimate. Recall that Methods 1-11 give the same group point estimate by means of the weighed arithmetic average given in Eq. (1). Methods 12 and 13 compute the point estimate by means of the geometric average given in Eq. (16). Table VIII gives the average squared error and associated standard error for both averages for each of the 27 simulation runs.

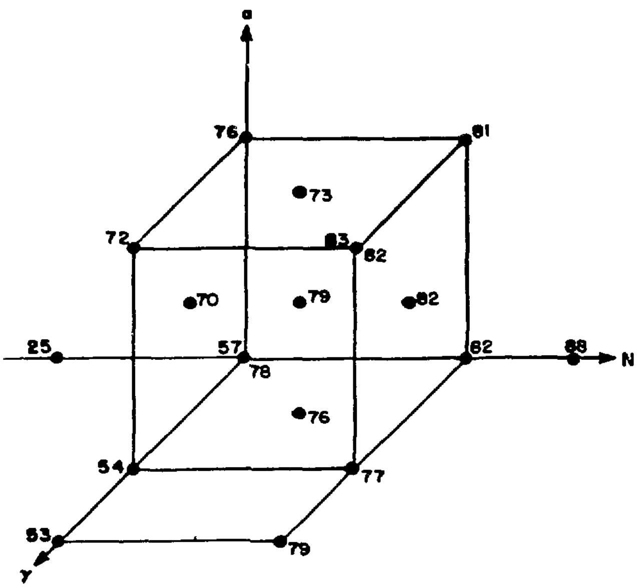

Fig. 4. Average coverage within the 90% theoretical interval for Method 2.

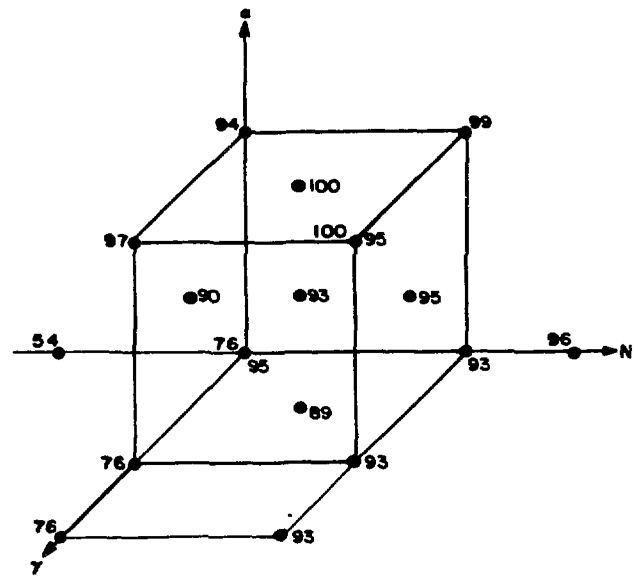

Fig. 5. Average coverage within the 90% theoretical interval for Method 3.

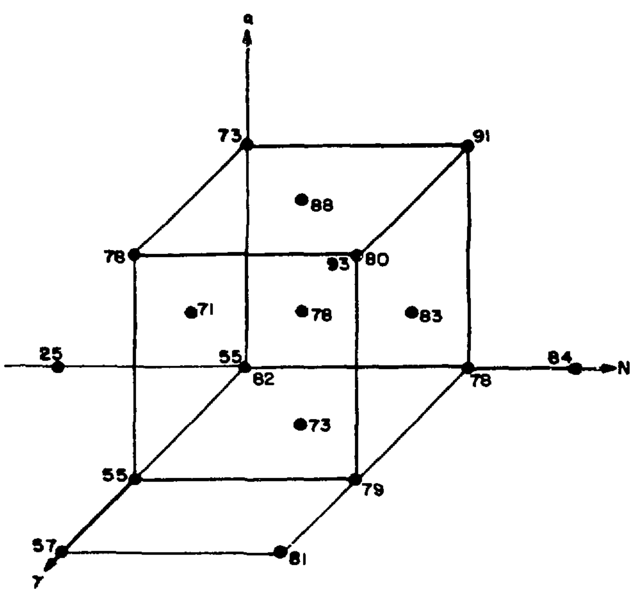

Fig. 6. Average coverage within the 90% theoretical interval for Method 4.

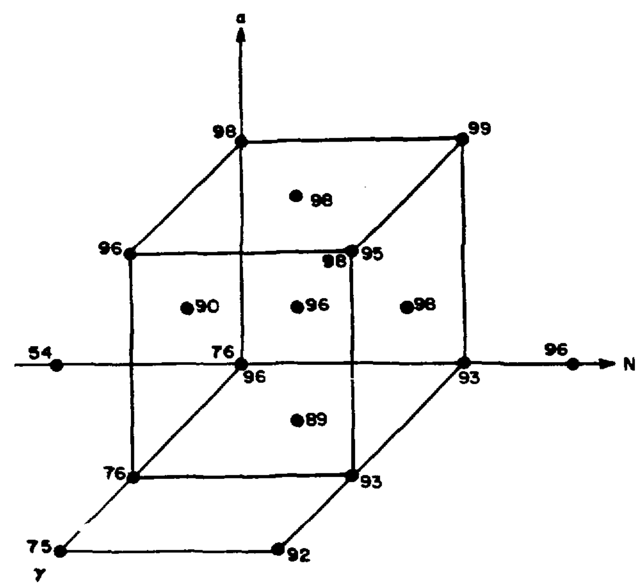

Fig. 7. Average coverage within the 90% theoretical interval for Method 5.

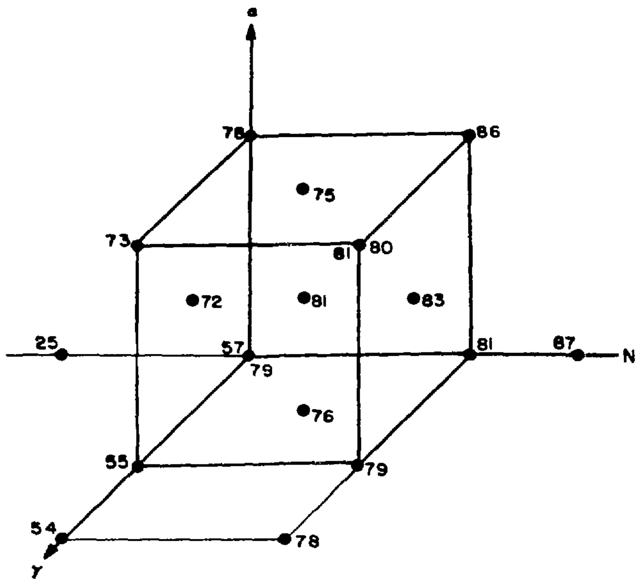

Fig. 8. Average coverage within the 90% theoretical interval for Method 6.

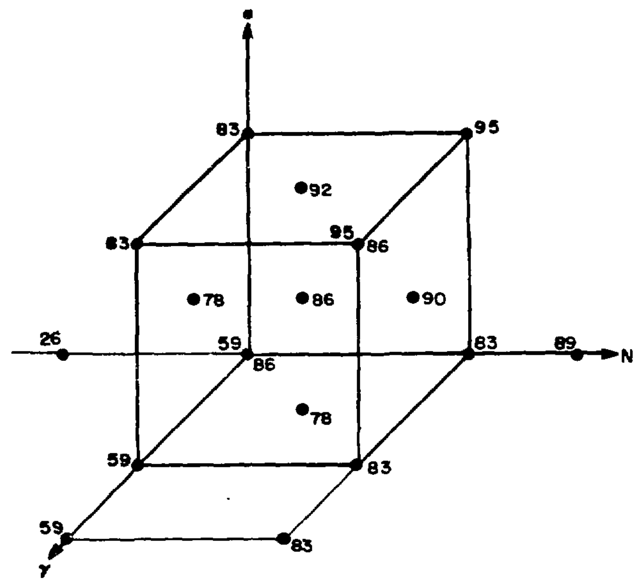

Fig. 9. Average coverage within the 90% theoretical interval for Method 7.

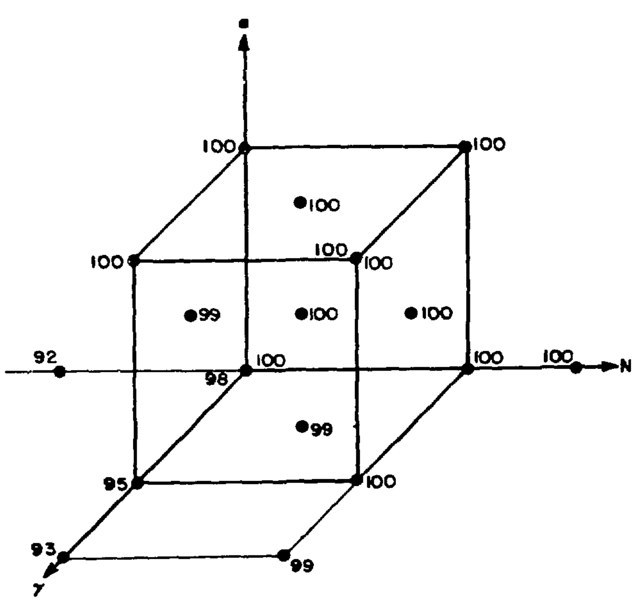

Fig. 10. Average coverage within the 90% theoretical interval for Method 8.

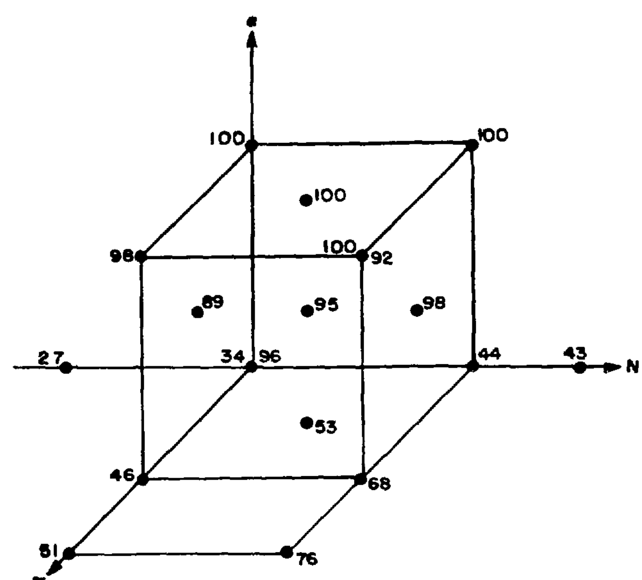

Fig. 11. Average coverage within the 90% theoretical interval for Method 9.

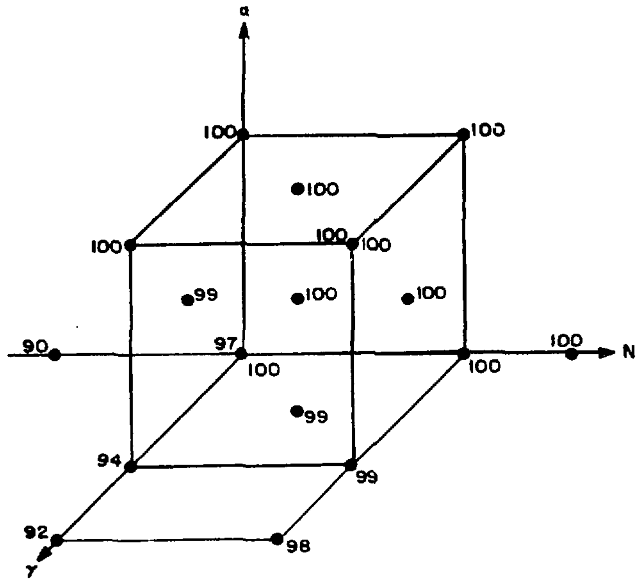

Fig. 12. Average coverage within the 90% theoretical interval for Method 10.

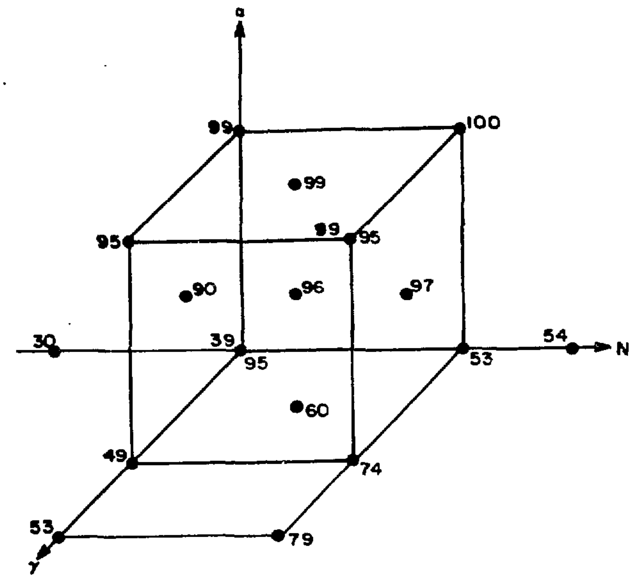

Fig. 13. Average coverage within the 90% theoretical interval for Method 11.

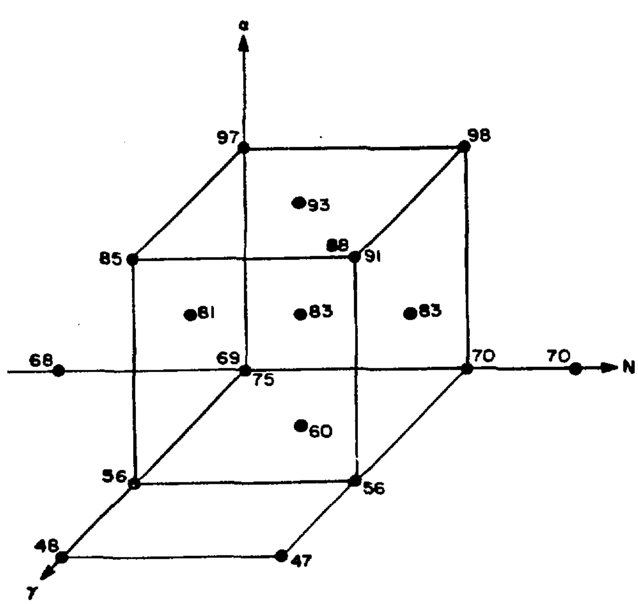

Fig. 14. Average coverage within the 90% theoretical interval for Method 12.

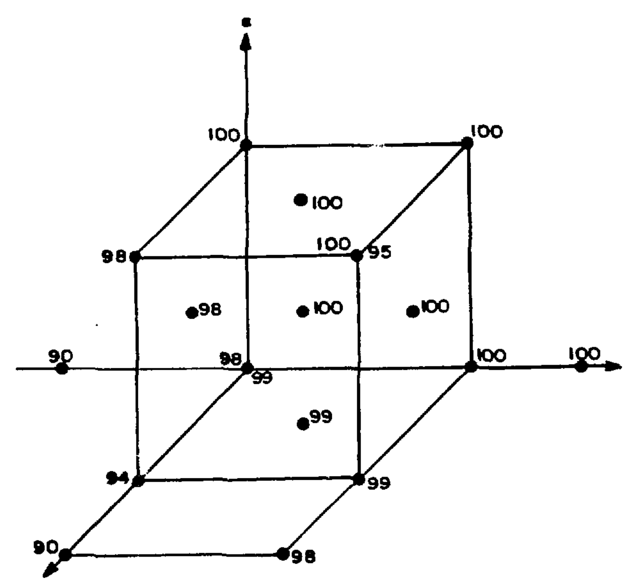

Fig. 15. Average coverage within the 90% theoretical interval for Method 13.

TABLE VII

AVERAGE AND AVERAGE SQUARED ERROR OF THE
COVERAGE PROBABILITY WITHIN THE
90 PERCENT INTERVAL FOR SIMULATION CASE NUMBER 1

Method	Average	STD Error	Average Squared Error	STD Error
1	7.62E-01	2.31E-02	7.24E-02	1.11E-02
2	5.66E-01	2.29E-02	1.62E-01	1.62E-02
3	7.60E-01	2.29E-02	7.22E-02	1.10E-02
4	5.51E-01	2.21E-02	1.71E-01	1.63E-02
5	7.63E-01	2.30E-02	7.16E-02	1.11E-02
6	5.71E-01	2.22E-02	1.58E-01	1.59E-02
7	5.86E-01	2.25E-02	1.49E-01	1.56E-02
8	9.81E-01	4.75E-03	8.88E-03	2.83E-04
9	3.38E-01	4.05E-02	4.80E-01	3.68E-02
10	9.74E-01	5.76E-03	8.85E-03	4.09E-04
11	3.89E-01	4.25E-02	4.42E-01	3.81E-02
12	6.91E-01	2.19E-02	9.18E-02	1.24E-02
13	9.82E-01	4.51E-03	8.68E-03	2.87E-04

V. CONCLUSIONS

The conclusions are based on the results of the simulation study described in the preceding section. First, let us consider the group point estimate. From Table VIII, the geometric average used in Methods 12 and 13 for combining the individual failure-rate estimates was observed to yield a smaller average squared error in 17 of the 27 cases. Although it is inconclusive here, it appears that geometric averaging is slightly preferred over arithmetic averaging as a method for combining individual failure-rate estimates into a group point estimate. Recall, that this method was the one finally selected for use in the IEEE Project 500.

Now consider the proposed 13 methods for combining the individual failure-rate estimates into a group interval estimate. Based on Figs. 3, 5, and 7, Methods 1, 3, and 5 provide interval estimates whose coverage probability exceeds 90% for most of the cases considered, particularly for large values of N and/or large values of α. For $N \leq 5$ and $\alpha = 0.25$, all three methods yielded interval estimates which, on the average, contained significantly less than 90% probability. For $n = 2$, only 54% coverage probability was observed. Generally, these three methods can be useful in providing somewhat conservative group interval estimates.

TABLE VIII

AVERAGE SQUARED ERROR FOR THE ARITHMETIC AND GEOMETRIC AVERAGES AS GROUP POINT ESTIMATES FOR THE 27 SIMULATION RUNS

Case Number	Average Squared Error		Standard Error	
	Arithmetic Average	Geometric Average	Arithmetic Average	Geometric Average
1	2.94×10^{-10}	3.45×10^{-11}	2.58×10^{-10}	3.14×10^{-13}
2	2.94×10^{-10}	3.45×10^{-11}	2.58×10^{-10}	3.14×10^{-13}
3	5.08×10^{-11}	3.53×10^{-11}	1.30×10^{-11}	1.06×10^{-13}
4	5.08×10^{-11}	3.53×10^{-11}	1.30×10^{-11}	1.06×10^{-13}
5	2.92×10^{-11}	3.46×10^{-11}	3.53×10^{-12}	2.91×10^{-13}
6	2.92×10^{-11}	3.46×10^{-11}	3.53×10^{-12}	2.91×10^{-13}
7	4.08×10^{-11}	3.54×10^{-11}	1.05×10^{-11}	8.95×10^{-14}
8	4.08×10^{-11}	3.54×10^{-11}	1.05×10^{-11}	8.95×10^{-14}
9	2.12×10^{-12}	1.93×10^{-12}	3.85×10^{-13}	2.77×10^{-13}
10	2.12×10^{-12}	1.93×10^{-12}	3.85×10^{-13}	2.77×10^{-13}
11	5.18×10^{-13}	6.31×10^{-13}	6.85×10^{-14}	7.07×10^{-14}
12	5.18×10^{-13}	6.31×10^{-13}	6.85×10^{-14}	7.07×10^{-14}
13	2.03×10^{-12}	1.85×10^{-12}	3.42×10^{-13}	2.72×10^{-13}
14	2.03×10^{-12}	1.85×10^{-12}	3.42×10^{-13}	2.72×10^{-13}
15	5.95×10^{-13}	7.67×10^{-13}	3.93×10^{-14}	4.03×10^{-14}
16	5.95×10^{-13}	7.67×10^{-13}	3.93×10^{-14}	4.03×10^{-14}
17	8.24×10^{-12}	1.59×10^{-11}	1.28×10^{-12}	7.02×10^{-13}
18	7.30×10^{-13}	8.40×10^{-13}	9.61×10^{-14}	1.08×10^{-13}
19	2.12×10^{-11}	1.75×10^{-11}	4.40×10^{-12}	9.29×10^{-13}
20	1.18×10^{-11}	1.54×10^{-11}	3.32×10^{-12}	8.28×10^{-13}
21	4.49×10^{-11}	3.50×10^{-11}	1.24×10^{-11}	2.06×10^{-13}
22	1.18×10^{-11}	1.54×10^{-11}	3.32×10^{-12}	8.28×10^{-13}
23	1.18×10^{-11}	1.54×10^{-11}	3.32×10^{-12}	8.28×10^{-13}
24	8.82×10^{-11}	3.58×10^{-11}	2.99×10^{-11}	2.54×10^{-14}
25	5.08×10^{-11}	3.53×10^{-11}	1.30×10^{-11}	1.06×10^{-13}
26	2.94×10^{-10}	3.45×10^{-11}	2.58×10^{-10}	3.14×10^{-13}
27	7.37×10^{-10}	3.44×10^{-11}	4.36×10^{-10}	2.46×10^{-12}

From Figs. 4, 6, and 8, Methods 2, 4, and 6 are observed to be significantly less conservative, yielding variable coverages between 25 and 88% for Method 2; 25 and 93% for Method 4; and 25 and 87% for Method 6. Again, the coverages tend to be quite low for $N \leq 5$ and $\alpha = 0.25$. Coverages tended to exceed 80% for $N \geq 15$, particularly when $\alpha = 10.0$.

By observation of Fig. 9, Method 7 provides somewhat variable average coverages. The coverages range from 26 to 95%. The coverages tend to be large for $N \geq 15$ and $\alpha = 10.0$, while they tend to be small for $N \leq 5$ and $\alpha = 0.25$.

On the other hand, Method 8 yields extremely conservative interval estimates, with most coverages exceeding 98%. From Fig. 11, Method 9 provides extremely variable coverages ranging from 27 to 100%. Based on Figs. 10 and 11,

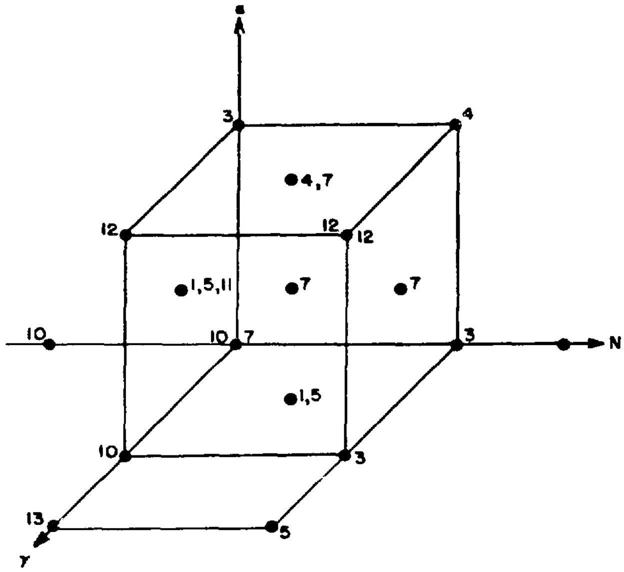

Fig. 16. The method having the coverage probability closest to 90%.

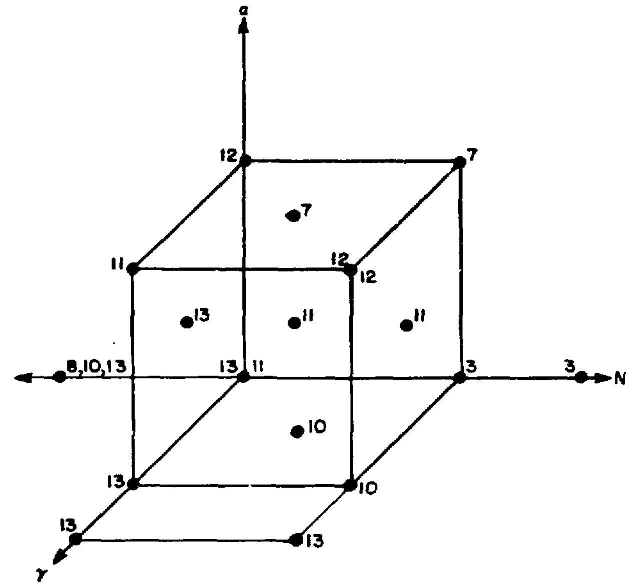

Fig. 17. The method having the minimum average squared error of the coverage probability within the 90% interval.

it appears that estimation of the prior variance ψ^2 by means of Eq. (14) is not a particularly good estimate to use.

Now consider Methods 10 and 11. Based upon Fig. 12, Method 10 yields extremely conservative interval estimates with most coverages in excess of 90%. From Fig. 13, Method 11 is less conservative than Method 10. Low coverages are observed for $\alpha = 0.25$ and all values of N, while high coverages are reported for $N \geq 10$ and $\alpha = 1.0$. In general, the coverages are quite variable ranging from 30 to 100% within the design.

Recall that Methods 12 and 13 were used in the IEEE Project 500. Method 12 generally provides coverages which underestimate the desired 90% value, except for $\alpha = 10.0$. The parameter γ appears to have a significant effect on the coverages with coverages decreasing as γ increases. In fact, for $\gamma = 0.20$, the smallest coverages, 47 and 48%, were observed. Coverages are generally unsatisfactorily small for $\alpha = 0.25$, regardless of the values of γ and N. Method 13 provides conservative coverages, most of which exceed 98% in the design. Method 13 is not as sensitive to γ as Method 12.

From Fig. 16, it is clearly observed that no method is preferred for the entire design. For example, for $N \leq 5$ and $\alpha = 0.25$, Method 10 is the preferred method. For $N \geq 10$ and $\alpha = 0.25$, Methods 3 and 5 are preferred. For $\alpha = 10.0$ and $\gamma = 0.10$, Method 12 is preferred. For $\alpha = 1.0$ and $N \geq 10$, Method 7 is superior.

Since the average does not describe the entire picture, let us consider the best average squared error performance methods presented in Fig. 17. Certain performance aspects are readily apparent. For $N \leq 5$, the simple interval estimates proposed in Methods 12 and 13 are superior. For $N \geq 15$, $\gamma = 0.01$, and $\alpha = 0.25$, the more complicated interval estimation method given in Method 3 is superior. For $\gamma = 0.20$, Method 13 is superior. Method 11 is superior for $\alpha = 1.0$, except for $N = 5$.

Some general conclusions can now be made. Again, it is stressed that these are based only on the results of the simulation study conducted here. Generally, the best method to use depends upon the value of the prior gamma shape parameter α, the parameter γ which controls the coverage of the individual interval estimates, and the number of data sources N which are being combined. It is clear that simple averaging methods (such as Methods 12 and 13) are as good or better than more complicated methods (such as Methods 1-6) when $N \leq 5$. More complicated methods should be considered for $5 < N \leq 15$, and are likely to yield superior group interval estimates when $N \geq 25$. Methods 2, 6, and 9, which are all based on the use of a gamma prior distribution, were not observed to be superior for any point in the design, and should likely be eliminated from further consideration. On the other hand, Methods 3, 7, 10, 11, 12, and 13 deserve serious consideration as useful methods for pooling failure-rate data.

REFERENCES

1. A. Briepohl, "Group Reliability Predictions," to appear in IEEE Transactions on Reliability, 1978.

2. IEEE Guide to the Collection and Presentation of Electrical, Electronic, and Sensing Component Reliability Data for Nuclear-Power Generating Station, IEEE Std 500-1977, The Institute of Electrical and Electronic Engineers, Inc., 1977.

3. H. F. Martz, "Pooling Life-Test Data by Means of the Empirical Bayes Method," IEEE Transactions on Reliability, Vol. R-24, No. 1, April 1975, pp. 27-30.

4. R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics, (3rd Edition), London: The Macmillan Company, 1970.