Probability and Statistics Notes

Note 3

The Assessment of System and Component Reliabilities Based on Both System and Component Test Results

Robert G. Easterling
Statistics and Computing Division 1643
Sandia Laboratories, Albuquerque 87115

December 1970

Abstract

Two problems are considered in this report: the determination of confidence limits for system reliability based on component results with and without the inclusion of system results, and the determination of confidence limits for component reliability based on component and system results. The method of maximum likelihood is used to obtain estimates of the first two moments of the maximum likelihood estimate of reliability. These estimates are then equated to the estimates of the moments under binomial sampling theory to obtain pseudo-sample size and number of successes from which confidence limits are computed. Some examples are given.

ACKNOWLEDGMENT

I would like to express my appreciation to H. E. Anderson, 1643, who carried out the computing work required in this report. I am also indebted to various neo-Bayesians, who shall go unnamed, for motivating this investigation, and to members of 1640 and 7425, for helpful discussion.

CONTENTS

TABLES

THE ASSESSMENT OF SYSTEM AND COMPONENT RELIABILITIES BASED ON BOTH SYSTEM AND COMPONENT TEST RESULTS

Introduction

Consider a system which is comprised of m components, each of which, when subjected to some test, has probability of success (reliability) p_i, and suppose that the system reliability under some test, is given by a function of the p_i's, say $h(p_1, p_2, \ldots, p_m)$, which we abbreviate by h(p). Suppose further that results are available from n_i tests on the ith component and (possibly) n tests on the system. The problems considered in this report are:

1. The determination of a 100γ-percent confidence interval on h(p) based only on component tests,

2. The determination of a 100γ-percent confidence interval on h(p) based on component and system tests,

3. The determination of a 100γ-percent confidence interval on p_i based on component and system tests.

The first problem is the one most often encountered in practice and which has attracted the most interest (see references [1], [2], and [4]-[8]). The determination of exact confidence limits for all but the simplest of systems is prohibitively complex, so the need is apparent for approximate confidence limits. Myhre and Saunders [6] have compared two approaches to obtaining approximate limits, and a third (first proposed by the author in [2]) is given in this report. Comparisons of the three approaches in [2] are repeated here. This third method leads directly to a solution to the second problem. The third problem has been considered from the standpoint of point estimation by Easterling and Prairie [3] for the special case of identical components, either in series or parallel, and the approach presented there is re-examined here from the standpoint of interval estimation. Some additional results pertaining to non-identical components are also given.

The implications of the assumption that the system reliability is h(p) should be emphasized. This requires that the reliability of a component be the same under the component test conditions as under the system test conditions and that the system reliability be capable of being expressed as a function of the component reliabilities. Inherent in this is the assumption of no interaction between components, that is, the success or failure of one component in a system test does not affect the performance of the other components. If the failure of one component increases the stress on another and degrades its reliability, then the assumption that the p_i's are the same in both tests is not met. Even if the condition of no interaction can be justified, the assumption of the model further requires that care should be given in designing component and system tests so that they are compatible. For example, the results of component tests at extreme conditions would not be applicable to the assessment of system reliability under moderate conditions, and so if the goal of the testing program is the latter, the components would have properly been used in a test at moderate conditions. However, even in cases where the assumption is not strictly met, system reliability is often estimated through a function such as h (p). An accompanying confidence limit statement is valuable in that it reflects the statistical precision in that estimate.

Confidence Intervals for System Reliability

The general procedure for obtaining a 100γ-percent confidence interval for a given parameter is to determine those values of the parameter for which the probability of obtaining a value of the random variable as extreme or more so than that observed is $1 - \gamma$. That is, values of the parameter are determined for which the observation is included in a range of outcomes which has probability γ. To do this requires an ordering of the possible outcomes. The difficulty in applying this to the problem at hand is that the observation consists of a vector of component results for which the problem of ordering quickly becomes quite complex (see Steck [8]), and the question of the best ordering also arises. Thus, the need arises for a tractable method to obtain approximate confidence intervals for complex systems. Three have been proposed. One due to Madansky [5], is based on the asymptotic chi-square distribution of the logarithm of the likelihood ratio statistic. Another, due to Rosenblatt [7], is based on the asymptotic normality of a U-statistic, in this situation U being an unbiased estimate of the system reliability which is referred to as the "simulation" estimate. The third approximation is based on the asymptotic normality of maximum likelihood estimators, a method referred to by Madansky [5] as "linearization." In many cases the latter two methods are equivalent, in fact Myhre and Saunders [6] regard them as one method in their comparison of approximate confidence limits for system reliability; however, they are different in situations in which the system reliability is not a linear function of the component reliabilities, that is, in situations where the maximum likelihood estimate is not unbiased.

In [6], the authors compare the likelihood ratio (LR) method and the maximum likelihood (ML) method in several situations for which exact confidence limits can be obtained. Their comparisons indicate that the former method is the more accurate, however, it does require more computation. The purpose of this paper is to propose a modification of the maximum likelihood (MML) method which will improve the approximation without complicating the computation required. (The same modification can be made on the U-approximation, but because of the similarity, we will not consider this in detail.)

The Maximum Likelihood Method and a Modification

Let $h(p)$ be the system reliability function, where $p = (p_1, p_2, \ldots, p_m)$ is the vector of component reliabilities. If we have n_i tests on the ith component with x_i successes, then the maximum likelihood estimate of the system reliability is $h(\hat{p})$, where $\hat{p}_i = x_i/n_i$, and, by the theory of maximum likelihood estimation, $h(\hat{p})$ has an asymptotic variance of

$$\sigma^2 = \sum_{i=1}^{m} \left[\frac{\partial h(p)}{\partial p_i}\right]^2 \frac{p_i(1 - p_i)}{n_i} \quad . \tag{1}$$

Furthermore, $h(\hat{p})$ is asymptotically normally distributed, so that approximate confidence limits on $h(p)$ can be obtained by replacing the p_i in Equation (1) by $\hat{p}_i$ and treating $\left(h(\hat{p}) - h(p)\right)/\hat{\sigma}$ as though it had a standard normal distribution.

An undesirable property of this method is that it can lead to confidence limits which fall outside the unit interval, a result due, of course, to the fact that the bounded, possibly unsymmetric distribution of $h(\hat{p})$ is being approximated by an unbounded, symmetric one. This forced symmetry is also one reason the ML approximation fares poorly in comparison with the LR. Because $h(\hat{p})$ is bounded between zero and one and because it arises from binomial sampling, a procedure which has intuitive appeal is to treat $h(\hat{p})$ as the usual binomial estimate based on n trials. The value of n, which might be called the pseudo-sample size, is unknown, but for the purposes of approximation, we will estimate it from

$$\hat{\sigma}^2 = \frac{h(\hat{p})\left(1 - h(\hat{p})\right)}{\hat{n}},$$

that is by equating the estimated sampling variance of $h(\hat{p})$ under maximum likelihood theory to what it would be under binomial theory. Then the component test results can be regarded as being equivalent to system results of $\hat{n}$ tests with $\hat{x} = h(\hat{p})\hat{n}$ successes. In most cases, $\hat{n}$ and $\hat{x}$ will not be integers. However, since our intent here is to find a procedure which gives approximate confidence limits, this should not cause any conceptual difficulties.

In ordinary binomial sampling with x successes in n trials, a lower 100γ-percent confidence limit on the reliability is given by the solution for p_L in

$$I(p_L, x, n - x + 1) = 1 - \gamma,$$

where

$$I(s, \alpha, \beta) = \frac{\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^s t^{\alpha-1} (1 - t)^{\beta-1} dt,$$

the incomplete beta function. By analogy then, an approximate lower 100γ-percent confidence limit on $h(p)$ would be given by the solution for h_L in

$$I(h_L, \hat{x}, \hat{n} - \hat{x} + 1) = 1 - \gamma.$$

Similarly, an approximate upper confidence limit would be given by the solution for h_U in

$$1 - I(h_U, \hat{x} + 1, \hat{n} - \hat{x}) = 1 - \gamma .$$

This approximation has the advantage that it takes into account the asymmetry and boundedness of $h(\hat{p})$ and, moreover, it is exact when the system is reduced to one component. It also is intuitively appealing since it treats a binomial-like random variable as one. Computationally, it requires tables or a computer program of the incomplete beta function. The method fails in the case of all successes or failures as do the ML and LR methods; however, other arguments can be used to handle those extreme cases.

In order to compare this approximation with the ML and LR approximations and also to assess its accuracy, we will consider the situations given in [6].

Comparison with Exact Intervals

Consider first a system which operates successfully if k out of m identical components operate. Then, with the notation, $q_1 = 1 - p_1$,

$$h(p_1) = \sum_{i=k}^{m} \binom{m}{i} p_1^i q_1^{m-i}$$

$$= I(p_1, k, m - k + 1) .$$

Exact confidence limits, p_L and p_U, can be obtained for p_1 on the basis of n_1 tests with x_1 successes. Then, since $h(p_1)$ is a monotonic function of p_1, exact confidence limits on $h(p_1)$ are given by $h(p_L)$ and $h(p_U)$. The asymptotic variance of $h(\hat{p}_1)$ is

$$\sigma^2 = \left[k\binom{m}{k} p_1^{k-1} q_1^{m-k}\right]^2 \frac{p_1 q_1}{n_1} .$$

From this, we can obtain $\hat{n}$ and $\hat{x}$ and compute confidence limits as outlined above. For example, with $n_1 = 25$, $x_1 = 23$, $m = 10$, and $\gamma = 0.975$, the MML approximation yields the two-sided 95-percent confidence intervals in Table I. In particular, with $k = 10$, $h(\hat{p}) = \hat{p}_1^{10} = 0.434$ and $\hat{\sigma}^2 = 100\hat{p}^{19}(1 - \hat{p})/25 = 0.0656$. Then

$\hat{n} = h(\hat{p})\left(1 - h(\hat{p})\right)/\hat{\sigma}^2 = 3.744$, $\hat{x} = 1.625$, and evaluating the incomplete beta leads to the pair of entries in Table I. The corresponding exact, maximum likelihood, and likelihood ratio limits, obtained from [6], are also given. The modified maximum likelihood limits compare favorably with those obtained by ML and by LR.

TABLE I

Comparison of 95-Percent Confidence Intervals
for a k Out of m System

k	Exact	MML	ML	LR
10	(0.049, 0.787)	(0.039, 0.915)	(-0.070, 0.938)	(0.076, 0.871)
9	(0.221, 0.977)	(0.218, 0.999)	(0.418, 1.21)	(0.2997, 0.9912)
8	(0.495, 0.9984)	(0.570, 1.000)	(0.823, 1.097)	(0.595, 0.9996)
7	(0.7511, 0.99993)	(0.868, 1.000)	(0.9664, 1.022)	(0.8266, 0.99994)

Consider now a serial system consisting of either two or three components with an equal number of tests, say n_o, on each component. Exact confidence limits in this situation are given in the tables by Lipow and Riley [4]. The MML approximate confidence limits follow from $h(p) = \prod_i^m p_i$ and

$$\sigma^2 = \prod_{j=1}^{m} p_j^2 \sum_{i=1}^{m} \frac{(1 - p_i)}{p_i n_o} .$$

Table II gives lower 100γ-percent confidence limits for the same cases given in [6]. One result of note in applying the MML method is that the resulting limits tend to be conservative. In order to alleviate this, $\hat{x}$ and $\hat{n}$ were rounded up to the nearest integer and confidence limits were obtained using these results; that is, for the purpose of obtaining an approximate confidence limit on system reliability we regard the component test results as being equivalent to $[\hat{n} + 1]$ system tests with $[\hat{x} + 1]$ successes where $[z]$ indicates the greatest integer less than z. The limits obtained in this way are labeled MMLI, I for integer. In general, it appears that both

TABLE II

Lower Confidence Limits for Series Systems
of Two or Three Components

	No. of Failures			Confidence Level									
				0.90					0.95				
m = 2	y_1	y_2		Exact	MML	MMLI	ML	LR	Exact	MML	MMLI	ML	LR
n_o = 10	1	1		0.607	0.570	0.585	0.655	0.629	0.548	0.514	0.530	0.611	0.571
	1	2		0.497	0.479	0.489	0.545	0.529	0.443	0.425	0.436	0.495	0.473
	2	2		0.445	0.407	0.441	0.456	0.451	0.392	0.356	0.391	0.405	0.397
	1	4		0.344	0.312	0.318	0.347	0.350	0.298	0.266	0.271	0.292	0.301
	2	3		0.364	0.337	0.362	0.373	0.375	0.304	0.290	0.315	0.320	0.326
n_o = 20	1	2		0.716	0.705	0.709	0.756	0.739	0.677	0.667	0.671	0.728	0.700
	2	2		0.683	0.656	0.669	0.701	0.687	0.643	0.617	0.631	0.670	0.647
	1	3		0.660	0.652	0.655	0.697	0.683	0.620	0.612	0.616	0.665	0.643
	2	3		0.622	0.608	0.619	0.647	0.636	0.582	0.568	0.580	0.614	0.597
	3	3		0.585	0.564	0.570	0.599	0.591	0.544	0.525	0.532	0.565	0.551
m = 3	y_1	y_2	y_3										
n_o = 20	1	1	1	0.747	0.710	0.721	0.760	0.743	0.709	0.671	0.684	0.732	0.705
	1	1	2	0.693	0.660	0.669	0.704	0.690	0.644	0.621	0.631	0.673	0.651
	1	2	2	0.639	0.614	0.619	0.654	0.643	0.598	0.576	0.580	0.621	0.604
	1	2	3	0.595	0.570	0.587	0.605	0.596	0.544	0.531	0.549	0.571	0.557
n_o = 30	1	2	3	0.705	0.694	0.669	0.723	0.714	0.674	0.663	0.638	0.698	0.683
	1	1	1	0.825	0.797	0.803	0.835	0.822	0.796	0.769	0.775	0.816	0.794
	2	2	2	0.712	0.695	0.703	0.725	0.715	0.681	0.664	0.672	0.700	0.685
n_o = 50	1	2	4	0.789	0.784	0.788	0.805	0.798	0.767	0.762	0.766	0.788	0.776
	1	1	2	0.861	0.850	0.852	0.874	0.865	0.841	0.830	0.833	0.860	0.845
n_o = 100	1	1	2	0.929	0.923	0.923	0.936	0.931	0.918	0.912	0.913	0.929	0.920
	2	3	5	0.858	0.855	0.856	0.866	0.861	0.844	0.841	0.842	0.855	0.848

the MML and MMLI methods are better than the ML and moreover the MMLI seems to be as accurate as the LR. This latter result is particularly satisfying since the MMLI limits are considerably the easier to obtain. The require only tables of binominal confidence limits. For series systems, then, it appears that the approximation which is the most convenient also gives adequately accurate results. One anomaly occurs for $n_o = 30$, $(y_1, y_2, y_3) = (1, 2, 3)$. In this case the MMLI limit is considerably less than the MML. This is because $\hat{n} = 32.005$, $\hat{x} = 25.988$, essentially 26 successes in 32 trials. However, the MMLI approximation treats this as 26 out of 33. Thus, there are situations in which rounding up would seem to be inadvisable. These should be fairly evident, as in this case, so we will not attempt to detail a set of rounding rules.

Distributional Properties

That an approximate method gives results which are generally in good agreement with those obtained by an exact method, is a desirable property. However, of more practical interest is the question of how well the approximate method works in situations for which the exact method cannot be used. In particular, does the method lead to confidence intervals which contain the system reliability with approximately the desired frequency?

To investigate this, we first consider, by simulation, the 2 out of 3 quorum system considered by Myhre and Saunders [6]. For this system, with nonidentical components, $h(p) = p_3\big(1 - (1 - p_1)(1 - p_2)\big) + (1 - p_3)p_1p_2$ and

$$\sigma^2 = \sum_{i=1}^{3} (p_j + p_k - 2p_jp_k)^2 p_i(1 - p_i)/n_i , \quad i \neq j,\ j \neq k,\ i \neq k .$$

For the simulation, each p_i was taken to be 0.7, for which $h(p) = 0.784$. Samples of size n_i were generated for each of the components and the MML and MMLI lower confidence limits obtained. This was repeated 2000 times and the proportion of times for which the lower limits were less than $h(p) = 0.784$ was determined. The results are given in Table III. The corresponding ML and LR frequencies were determined as nearly as possible from the plots given in [6]. The MML and MMLI limits seem to fall between the ML and LR in accuracy, similar to the results of the previous section, with the MMLI slightly more accurate. The one exception which stands out, $n_o = 10$, confidence level = 0.95, indicates that rounding to the next integer may not be advisable with samples this small.

TABLE III

Simulation Results for 2 Out of 3 Quorum System

$p_1 = p_2 = p_3 = 0.7$

$n_1 = n_2 = n_3 = n_o$

	Confidence Level							
	0.90				0.95			
n_o	MML	MMLI	ML	LR	MML	MMLI	ML	LR
10	0.908	0.899	0.80	0.92	0.955	0.912	0.92	0.95
20	0.923	0.889	0.87	0.90	0.960	0.967	0.92	0.95

In further investigation of the properties of the MML and MMLI methods, three simple two-component systems for which the confidence levels could be determined exactly were considered. These systems, their reliabilities and asymptotic variances of $h(\tilde{p})$ are:

1. Series

$$h(p) = p_1 p_2 ,$$

$$\sigma^2 = p_1^2 p_2 (1 - p_2)/n_2 + p_2^2 p_1 (1 - p_1)/n_1 .$$

2. Parallel

$$h(p) = 1 - (1 - p_1)(1 - p_2) ,$$

$$\sigma^2 = (1 - p_1)^2 p_2 (1 - p_2)/n_2 + (1 - p_2)^2 p_1 (1 - p_1)/n_1 .$$

3. Series - Parallel

$$h(p) = p_1\left(1 - (1 - p_2)^2\right) ,$$

$$\sigma^2 = \left(1 - (1 - p_2)^2\right)^2 p_1(1 - p_1)/n_1 + 4p_1^2 p_2(1 - p_2)^3/n_2 .$$

The cases considered and the corresponding confidence levels are given in Table IV. The component reliabilities were chosen so that h(p) = 0.8 in all cases. The actual confidence levels were determined by enumerating those cases for which the lower confidence limit exceeded 0.8, accumulating the probabilities of those cases, and then taking the complement of this. For the case of no failures, the equivalent sample size was taken to be min (n_1, n_2) for the first and third systems, and max (n_1, n_2) for the parallel system. For the most part the MML and MMLI intervals are conservative, the latter being slightly less so. However, so are ordinary binomial confidence limits because of the discreteness of the random variable. Since the MML and MMLI approximations are an adaptation of this approach, this result then is not surprising. In fact, if 20 system tests were run for which the system reliability is 0.8, lower 95-percent binomial confidence limits would contain 0.8 with probability 0.980, and lower 90-percent confidence limits would contain 0.8 with probability 0.931. The corresponding probabilities for 30 system tests are 0.956 for both confidence levels; with 50 system tests, both 90- and 95-percent lower confidence limits are less than 0.8 with probability 0.952. Thus, the distributional properties of the approximate confidence limits based on component results compare very well with what would be obtained from system tests, and in some cases, notably a parallel system, the limits are actually more accurate in terms of the actual confidence level obtained.

The results presented here are far from conclusive. However, they do indicate that the approach presented here gives system confidence limits which adequately approximate exact confidence limits and also have distributional properties similar to those of confidence limits based on system tests. There seems to be little more than one could ask of an approximation. The computation required is not too complicated, involving only the maximum likelihood estimation of the asymptotic variance of the maximum likelihood estimate of system reliability and evaluation of the incomplete beta function. The latter can be facilitated and the approximation improved by rounding the derived pseudo-successes and sample size to the next integer.

One situation in which the method may give inaccurate results is when a component has no failures, since in this case, the estimated variance, $\hat{p}_i(1 - \hat{p}_i)/n_i$, is zero, so the method treats this as a perfect component. As an extreme case, suppose we have a two component serial system and suppose the first component has been tested once with a success and the second has 99 successes out of 100 tests. Then the method treats these results as being equivalent to 99 successes in 100 system tests, which is obviously optimistic. One solution would be to let $\hat{n} = n_{min} = 1$ and treat the results as $\hat{x} = 0.99$, $\hat{n} = 1$. However, this disregards the variability in the second component reliability estimate. Another approach would be to estimate p_i in $p_i(1 - p_i)/n_i$ by a lower 50 percent confidence limit in those cases for which $\hat{p}_i = 1$, so that the estimated variance would not be zero. The asymptotic variance (1) would still hold since all that is required is a consistent estimate of the variance of $\hat{p}_i$. Many other techniques could be used to avoid a zero estimate of the variance of an estimated component reliability; however, there seems to be little basis for a choice other than the user's preference. The essence of the approach presented here is the way in which $h(\hat{p})$ and the estimated variance are used to get approximate confidence limits, not the way in which the variance is estimated. In cases where at least one failure and one success have been obtained on all components, or if no failures have been observed on a component and that component has been tested at least as many times as any other component in the system, this approach, using the maximum likelihood estimate of σ^2, would seem to give accurate confidence limits.

This method also conveniently admits the inclusion of system test results, assuming that they are compatible, since it essentially amounts to approximating the likelihood function of h(p) by the binomial likelihood function with arguments $\hat{x}$ and $\hat{n} - \hat{x}$, that is, $h(p)^{\hat{x}}\left(1 - h(p)\right)^{\hat{n}-\hat{x}}$. The likelihood of h(p) based on n system tests with x successes is this same function with arguments x and n - x, so the joint likelihood based on component and system tests is $h(p)^{x+\hat{x}}\left(1 - h(p)\right)^{\hat{n}+n-x-\hat{x}}$. Thus, for the purpose of obtaining system confidence limits on h(p), the combined results are equivalent to $n + \hat{n}$ tests with $x + \hat{x}$ successes and either the MML or MMLI methods can be used directly. The pseudo system results $(\hat{x}, \hat{n})$ and the

TABLE IV

Actual Confidence Level Attained Using MML and MMLI Methods: Two Component Systems

System	n_1	n_2	Intended Confidence Level 0.90 MML	0.90 MMLI	0.95 MML	0.95 MMLI
Series						
$p_1 = 0.9$, $p_2 = 0.889$	20	20	0.934	0.934	0.988	0.988
	30	20	0.951	0.951	0.961	0.961
	30	30	0.959	0.959	0.959	0.959
	50	50	0.915	0.913	0.960	0.960
Parallel						
$p_1 = 0.6$, $p_2 = 0.5$	20	20	0.915	0.901	0.959	0.951
	30	20	0.917	0.914	0.965	0.949
	30	30	0.917	0.918	0.962	0.952
	50	50	0.917	0.914	0.958	0.953
Series - Parallel						
$p_1 = 0.9$, $p_2 = 0.667$	20	20	0.948	0.948	0.892	0.966
	30	20	0.952	0.952	0.984	0.971
	30	30	0.950	0.950	0.975	0.964
	50	50	0.928	0.931	0.968	0.960

system results (x, n) can be statistically compared by using standard statistical techniques, such as a χ^2 test, for comparing binomial test results. This could provide an indication as to whether the system reliability is consistent with the model, h(p); that is, the assumption of compatibility can be tested before the results are pooled. This could be quite valuable in identifying problems in the system due to considerations other than component failures.

Confidence Intervals for Component Reliability

Identical Components

Suppose now that the objective is to incorporate the results of system tests with those of component tests in order to improve the assessment of component reliability; and as a starting point, let us only consider simple systems of m identical, independent components in series or parallel. The justification for wanting to improve the assessment of component reliability is that some future system may require this component, and so in order to assess the reliability of the new system it is desirable to use as much component information as possible. A specific example of this is the bridgewire used in a certain thermal battery. This battery uses two bridgewires for activation, two being used strictly for the purpose of redundancy. When a battery is tested, both bridgewires are pulsed, and if the battery fires there is no way to determine whether one or both of the bridgewires functioned. If one bridgewire does not function, it in no way affects the performance of the other as there is no change in stress on the other. Over a period of time, data have been collected at the system level (battery tests) and at the component level (bridgewire tests). Because this type of bridgewire has other applications (in gas generators and explosive switches, for example) an estimate of the bridgewire reliability was desired using both battery and individual bridgewire results. In another situation encountered, the component was a section of cable and the system consisted of m sections of cable connected serially. Since a future application required only a single section, it was desired to combine the available data on systems and components to improve the component assessment. A related problem is determining an equivalence relation between the number of system tests and component tests where the component reliability is of concern. For example, how many system tests must be performed to obtain an estimate of component reliability that is as "good" as the estimate obtained from n individual component tests?

Consider first a system consisting of m components in series. Then the reliability of the system is p^m. For convenience, we have dropped the subscript on the component reliability. If we have n_1 component tests with x_1 successes and n system tests with x successes, then the respective maximum likelihood estimates of p are x_1/n_1 and $(x/n)^{1/m}$. A common procedure which suggests itself here would be to seek a linear combination of these two estimates which would be optimal in some sense. However, this is not feasible in this case, since a linear combination will not retain the maximum likelihood properties and since the mean and variance of $(x/n)^{1/m}$ can at best only be approximated. A procedure which can be usefully employed is that of maximum likelihood. The likelihood of the combined results is

$$L = \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} \binom{n}{x} p^{mx} \left(1 - p^m\right)^{n-x} ,$$

and the log likelihood, ignoring terms not involving p, is

$$\ell = (x_1 + mx)\, \ell n\, p + (n_1 - x_1)\, \ell n\, (1 - p) + (n - x)\, \ell n\, (1 - p^m) .$$

Differentiating this with respect to p and equating the result to zero, yields the following equation in $\hat{p}$, the maximum likelihood estimate of p.

$$(n_1 + mn)\hat{p}^m + (n_1 - x_1)(\hat{p}^{m-1} + \hat{p}^{m-2} + \ldots + \hat{p}) - (x_1 + mx) = 0 . \tag{2}$$

We note that this equation has a unique root between zero and $[(x_1 + mx)/(n_1 + mn)]^{1/m}$, since the left-hand side is negative for $\hat{p} = 0$, positive for $\hat{p} = 1$, and its derivative is positive throughout this range. The upper bound is obtained by deleting the second term which is always nonnegative. For the special case, $m = 2$, we obtain

$$\hat{p} = \frac{(x_1 - n_1) + \sqrt{(n_1 - x_1)^2 + 4(n_1 + 2n)(x_1 + 2x)}}{2(n_1 + 2n)} .$$

While the exact variance of $\hat{p}$ is unobtainable, the asymptotic variance is given by $-1/E\left((\partial^2 \ell)/(\partial p^2)\right)$, where E denotes expectation. Differentiating ℓ twice and taking the expectation yields

$$-E\, \frac{\partial^2 \ell}{\partial p^2} = \frac{n_1}{p} + mn + \frac{n_1}{1 - p} + \frac{mn(1 - p^2)(m - 1)p^{m-2}}{1 - p^m} + \frac{m^2 n(1 - p^2)p^{2m-2}}{\left(1 - p^m\right)^2} .$$

After some algebraic manipulations, we obtain var $(\hat{p}) \approx p(1 - p)/N(p)$, where

$$N(p) = n_1 + nm\left[\frac{mp^{m-1}(1 - p)}{1 - p^m}\right] . \tag{3}$$

Writing the asymptotic variance in this way indicates that N(p) can be regarded as the equivalent sample size. Thus, following the approach suggested by the previous section, we can obtain approximate confidence limits on p by considering $\hat{p}$ to be the usual binomial estimate of p based on a sample of size $\hat{n}_1 = N(\hat{p})$ with $\hat{x}_1 = \hat{p}N(\hat{p})$ successes. We again refer to these as MML limits, and to those obtained by rounding up as MMLI limits.

If the mn components used in the system test had been tested individually, then the total sample size for assessing component reliability would have been $n_1 + mn$. The term in Equation (3) by whicn mn is multiplied is less than one, so that testing components in series has resulted in less information being obtained on p compared to what would have been obtained if all the components had been tested singly. This term, as a function of p, is plotted in Figure 1.

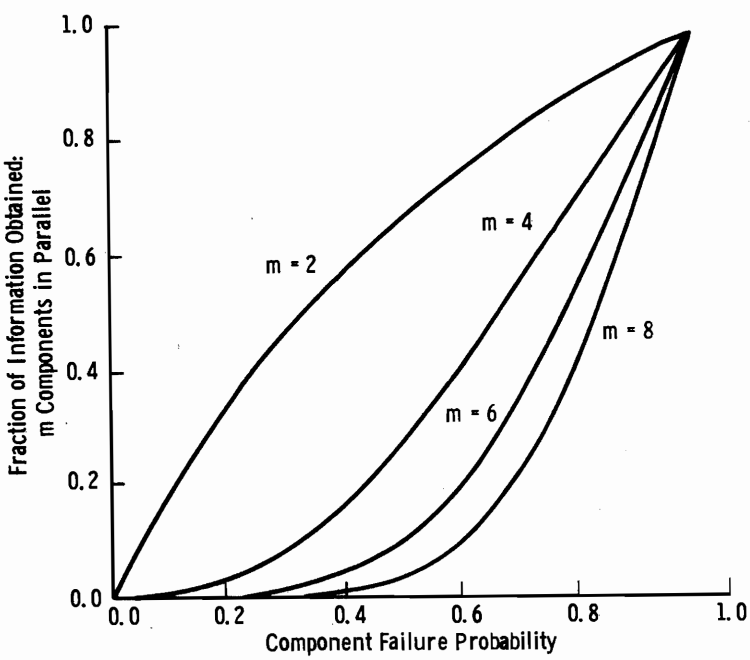


Figure 1.

As might be expected, it is an increasing function of p and a decreasing function of m for fixed m and p, respectively. The higher the reliability, the more information can be obtained by testing components in series, while the less reliable a component is, the less information can be obtained. This is because the most component information is obtained from the system test when a success occurs, since this means that m out of m components succeeded. (Failure of the system indicates only that 0, 1, 2, ..., or m-1 successes may have occurred.) Since the system reliability is an increasing function of p, it follows that the amount of component information is also an increasing function of p.

Suppose now that the system consists of m components in parallel. If the roles of failure and success are reversed, the failure probability for m components in parallel is analogous to the reliability of m components in series; that is, $(1 - p)^m$ in the former case, p^m in the latter. Thus, the solution for $\hat{p}$ in the series case can be obtained from Equation (2) with $1 - \hat{p}$, $n_1 - x_1$, and $n - x$ substituted for $\hat{p}$, x_1, and x, respectively. For the case of m = 2,

$$\hat{p} = \frac{2(n_1 + 2n) + x_1 - \sqrt{\left(2(n_1 + 2n) + x_1\right)^2 - 8(x_1 + x)(n_1 + 2n)}}{2(n_1 + 2n)} .$$

Additionally, the equivalent sample size and the fraction of the possible information obtained for the series case for a probability p of failure is the same as for the parallel case at 1 - p, so that this fraction can be obtained from Figure 1 by replacing p by 1 - p. For highly reliable components, more component information is obtained from series system tests than from parallel system tests. This is because a series system failure in this case is highly likely to be the result of only one component failure out of m trials, while parallel systems fail only if all m components fail, so that information from situations in which less than m failed is lost.

Distributional Properties

The results of a simulation in [3] indicated that the estimates given above for the component reliability and the variance of that estimate are not badly biased, even for quite small samples. Thus, it seems reasonable to expect the approximate confidence intervals to also be adequate. In an investigation of this, the actual confidence levels were calculated in a similar manner as in the previous section for two component series and parallel systems. The results are given in Table V.

TABLE V

Actual Confidence Levels for Component Reliability Based on Combined System and Component Results: Two Component Systems

System	n_1	n	N(p)	Desired Confidence Levels					
				0.90			0.95		
				MML	MMLI	Comp	MNL	MMLI	Comp
Series									
p = 0.9	20	10	38.9	0.918	0.918	0.920	0.985	0.985	0.985
	20	20	57.9	0.944	0.941	0.947	0.986	0.986	0.986
	40	10	58.9	0.946	0.946	0.947	0.986	0.986	0.986
	40	40	115.8	0.911	0.911	0.922	0.957	0.957	0.962
p = 0.8	20	10	37.8	0.912	0.912	0.924	0.968	0.965	0.972
	20	20	55.6	0.911	0.891	0.933	0.959	0.959	0.969
	40	10	57.8	0.923	0.923	0.933	0.965	0.965	0.969
	40	40	111.1	0.929	0.935	0.935	0.963	0.957	0.961
Parallel									
p = 0.9	20	10	23.6	1.000	1.000	0.920	1.000	1.000	0.985
	20	20	27.3	1.000	1.000	0.947	1.000	1.000	0.986
	40	10	43.6	0.923	0.923	0.947	0.987	0.987	0.986
	40	40	54.5	0.946	0.946	0.922	0.990	0.990	0.962
p = 0.8	20	10	26.7	0.954	0.954	0.924	0.992	0.954	0.972
	20	20	33.3	0.965	0.905	0.933	0.969	0.969	0.969
	40	10	46.7	0.947	0.942	0.933	0.979	0.947	0.969
	40	40	66.7	0.917	0.917	0.935	0.959	0.974	0.961

Also given are the actual confidence levels if all $n_1 + 2n$ components involved in the two tests had been tested singly. The MML and MMLI results are comparable to those that would have been obtained from all component tests, particularly for series systems. This is because more component information is obtained from series systems than for parallel systems for components with reliability as high as those considered, a fact which is reflected by the values of N(p) also given in the table. A test of 40 components, in which 20 are tested singly and the other 20 are tested in 10 two-component serial systems, is equivalent to 38.9 component tests, while if the systems are made up of two parallel components, the equivalent number of component tests is only 23.6. That is, in the first case, 10 system tests were equivalent to 18.9 component tests, but in the second case, to only 3.6

component tests. In situations such as the latter where little component information is gleaned from the system test, combining the data does not seem called for, since the possible introduction of bias is not compensated for by the slight gain in information.

Nonidentical Components

Suppose now that the system is comprised of nonidentical components. The maximum likelihood approach used above can be extended to more than one parameter as follows. Suppose that we have x_i successes in n_i tests of the ith component and x successes in n system tests for which the reliability is h(p). Then the log likelihood of the combined results is given by

$$\ell(p_1, p_2, \ldots, p_m) = \sum_{i=1}^{m} [x_i \ln p_i + (n_i - x_i) \ln (1 - p_i)]$$

$$+ x \ln h(p) + (n - x) \ln \left(1 - h(p)\right) .$$

The maximum likelihood estimates of the p_i's are obtained by solving the system of equations (called the normal equations)

$$\frac{\partial \ell}{\partial p_i} = 0 \qquad i = 1, 2, \ldots, m .$$

The asymptotic variances and covariances of the estimates are obtained by inverting an m x m matrix in which the i, jth element is given by

$$- E\left(\frac{\partial^2 \ell}{\partial p_i \partial p_j}\right) .$$

Approximate joint confidence regions for the p_i's, or a subset of them, can be obtained by applying multivariate normal distribution theory. Approximate confidence intervals for a single p_i can be obtained in a similar manner as before. Let ℓ^{ii} be the ith diagonal element of the variance-covariance matrix. Then the combined results yield an estimate of p_i, $\hat{p}_i$, and an estimated variance, $\hat{\ell}^{ii}$. Equating the estimated variance to $\hat{p}_i(1 - \hat{p}_i)/\hat{n}_i$, as before, then yields a pseudo-sample size and number of successes from which MML or MMLI confidence limits can be obtained.

Although the theory is a straightforward extension of that for identical components, the implementation is quite involved and tedious for all but the simplest systems, so it is not feasible unless a good deal of system information is available relative to the amount of component information. To illustrate, consider a system of two components in series. The log likelihood is

$$\ell(p_1, p_2) = x_1 \ln p_1 + (n_1 - x_1) \ln (1 - p_1) + x_2 \ln p_2 + (n_2 - x_2) \ln (1 - p_2)$$

$$+ x \ln p_1 p_2 + (n - x) \ln (1 - p_1 p_2) ,$$

and the normal equations are

$$\frac{x_1 + x}{\hat{p}_1} - \frac{(n_1 - x_1)}{1 - \hat{p}_1} - \frac{(n - x)\hat{p}_2}{1 - \hat{p}_1\hat{p}_2} = 0 ,$$

$$\frac{x_2 + x}{\hat{p}_1} - \frac{(n_2 - x_2)}{1 - \hat{p}_2} - \frac{(n - x)p_1}{1 - \hat{p}_1\hat{p}_2} = 0 .$$

An explicit solution to this pair of nonlinear equations cannot be obtained. However, the solution can be determined by using a technique such as the Newton-Raphson method for solving a system of nonlinear equations.

To get the asymptotic variances and covariance, we first obtain

$$- E\left(\frac{\partial^2 \ell}{\partial p_1^2}\right) = \frac{n_1 + np_2}{p_1} + \frac{n_1}{1 - p_1} + \frac{np_2^2}{1 - p_1 p_2} ,$$

$$-E\left(\frac{\partial^2 \ell}{\partial p_2^2}\right) = \frac{n_2 + np_1}{p_2} + \frac{n_2}{1 - p_2} + \frac{np_1^2}{1 - p_1 p_2} ,$$

$$- E\left(\frac{\partial^2 \ell}{\partial p_1 \partial p_2}\right) = \frac{n}{1 - p_1 p_2} .$$

Denoting these terms by ℓ_{11}, ℓ_{22}, and ℓ_{12}, the asymptotic variance covariance matrix of the maximum likelihood estimate of p_1 and p_2 is given by

$$V = \left(\ell_{11}\ell_{22} - \ell_{12}^2\right)^{-1} \begin{bmatrix} \ell_{22} & -\ell_{12} \\ -\ell_{12} & \ell_{11} \end{bmatrix}$$

$$= \begin{bmatrix} \ell^{11} & \ell^{12} \\ \ell^{12} & \ell^{22} \end{bmatrix} .$$

Then an approximate 100 γ-percent confidence interval on p_1 can be obtained as outlined above, that is by solving for $\hat{n}_i = \hat{p}_i(1 - \hat{p}_i)/\hat{\ell}^{ii}$, $\hat{x}_i = \hat{p}_i \hat{n}_f$ and proceeding as described for the MML or MMLI methods.

Examples

Confidence Limits for System Reliability

Consider first a two-component parallel system, for which $h(p) = 1 - (1 - p_1)(1 - p_2)$. Suppose that 100 tests have been performed on each component with 97 successes on the first, 95 on the second. Then, the maximum likelihood estimate of h(p) is $h(\hat{p}) = 1 - (0.03)(0.05) = 0.9985$. The asymptotic variance of $h(\hat{p})$ is

$$\sigma^2 = (1 - p_2)^2 \frac{p_1(1 - p_1)}{n_1} + (1 - p_1)^2 \frac{p_2(1 - p_2)}{n_2} ,$$

so

$$\hat{\sigma}^2 = 0.000001155 ,$$

$$\hat{n} = \frac{h(\hat{p})\left(1 - h(\hat{p})\right)}{\hat{\sigma}^2}$$

$$= 1296.75 ,$$

and

$$\hat{x} = 1294.81 .$$

An approximate lower 90-percent confidence limit on h(p), by the MML method is given by the solution for h_L in

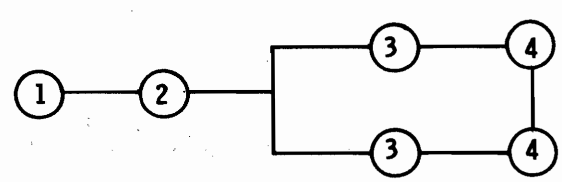
$$I(h_L, 1294.81, 2.94) = 0.10 ,$$

which is $h_L = 0.99596$. The MMLI lower limit is obtained from

$$I(h_L, 1295, 3) = 0.10 ,$$

which gives $h_L = 0.99590$. From [1], an exact 90-percent lower confidence limit on h(p) is 0.99588, and from [5] the LR limit is 0.99482 and the ML is 0.99836, so the techniques proposed here compare favorably. This example should also point out the fallacy in regarding the component results as being equivalent to the pseudo-system results. In no way is the assertion made that testing 100 components is equivalent to testing 1297 systems. It is merely claimed that for the purpose of calculating system confidence limits the component results can be treated in this way.

Consider now what may be a more representative system, one represented by the diagram opposite. The reliability of this system is given by



$$h(p) = p_1 p_2 \left(1 - (1 - p_3 p_4)^2\right) .$$

Suppose we have the following component test results,

i	n_i	x_i
1	20	20
2	40	39
3	50	48
4	50	49

from which we desire to obtain a 95-percent confidence limit on h(p). The first step is to obtain the asymptotic variance. The partial derivatives of h(p) with respect to the p_i are

$$\frac{\partial h(p)}{\partial p_1} = p_2\left(1 - (1 - p_3p_4)^2\right) ,$$

$$\frac{\partial h(p)}{\partial p_2} = p_1\left(1 - (1 - p_3p_4)^2\right) ,$$

$$\frac{\partial h(p)}{\partial p_3} = 2p_1p_2(1 - p_3p_4)p_4 ,$$

$$\frac{\partial h(p)}{\partial p_4} = 2p_1p_2(1 - p_3p_4)p_3 .$$

Thus, the asymptotic variance of $h(\hat{p})$ is

$$\sigma^2 = \sum_{i=1}^{4} \left[\frac{\partial h(p)}{\partial p_i}\right]^2 \frac{p_i(1 - p_i)}{n_i}$$

$$- \left(1 - (1 - p_3p_4)^2\right)^2 \left(p_2^2p_1(1 - p_1)/n_1 + p_1^2p_2(1 - p_2)/n_2\right)$$

$$+ 4p_1^2p_2^2(1 - p_3p_4)^2\left(p_4^2p_3(1 - p_3)/n_3 + p_3^2p_4(1 - p_4)/n_4\right) .$$

Replacing the p_i's by their maximum likelihood estimates leads to $h(\hat{p}) = 0.9716$ and $\hat{\sigma}^2 = 0.0006198$. The pseudo-sample size and number of successes are obtained from $\hat{n} = h(\hat{p})\left(1 - h(\hat{p})\right)/\hat{\sigma}^2 = 44.52$ and $\hat{x} = h(\hat{p})\hat{n} = 43.26$. The lower MML confidence limits, obtained by evaluating the incomplete beta function with parameters 43.26 and 2.26, are 0.907 and 0.889 for 90- and 95-percent confidence levels, respectively. (If only a rough answer is desired, these limits could have been obtained by interpolation on a confidence limit slide rule computer.) Rounding n and x up gives pseudo-system results of 1 failure in 45 trials, for which the corresponding confidence limits are 0.916 and 0.899. This is one case where direct use of the MML or MMLI method may be unduly optimistic since the maximum likelihood estimate of the variance of the reliability of the first component, on which there are but 20 tests, is zero. Estimating p_1 by its lower 50 percent confidence limit, which is 0.965, and substituting this into $p_1(1 - p_1)/n_1$ in the above equation for σ^2 leads to $\hat{n} = 12.432$, $\hat{x} = 12.079$ for which the lower MML 95 percent confidence limit is 0.742 and for the MMLI method is 0.794. These are considerably more conservative than those obtained above, but more accurately reflect the precision in the estimate, $h(\hat{p})$. However, this does not imply that direct use of the MML or MMLI methods would not have the confidence properties ascribed to them in repeated sampling. If there had been 50 tests of component 1 with no failures, this same approach leads to $\hat{n} = 31.309$, $\hat{x} = 30.420$ for which the MML 95 percent lower confidence limit is 0.863 and the MMLI is 0.860, not greatly different from those obtained ignoring the first component.

Suppose now that in addition to the above component results, 30 systems have been tested with 28 successes. Under the assumption that the reliability of the system is h(p), these results can be combined with those obtained above. This gives $\hat{n} = 74.52$, $\hat{x} = 71.26$, for which the MML and MMLI confidence limits are 0.908 and 0.913, respectively, for the 90-percent confidence level and 0.895 and 0.900 for 95 percent.

Confidence Limits for Component Reliability

Identical Components -- Consider first a series system of 2 identical components, and suppose we have the following results:

	No. of Tests	No. of Successes
Component	30	29
System	40	36

Then

$$\hat{p} = \frac{x_1 - n_1 + \sqrt{(n_1 - x_1)^2 + 4(n_1 + 2n)(x_1 + 2x)}}{2(n_1 + 2n)}$$

$$= 0.954 ,$$

$$N(\hat{p}) = n_1 + \frac{nm^2\hat{p}(1 - \hat{p})}{1 - \hat{p}^2} = 108.10 ,$$

and

$$\hat{x}_1 = 103.10 .$$

Based on this, lower 90- and 95-percent MML confidence limits on the component reliability are 0.916 and 0.905 and the corresponding MMLI limits are 0.917 and 0.906. Note that in this case a considerable amount of component information was obtained from the system test, since $N(\hat{p})$ is 108.10 and the total number of components used in the test is 110. This is because with components of this high a reliability, a system failure is most probably due to the failure of one but not both of the components so that the system results most likely were due to 4 failures in 80 components. Based on the component tests alone, 90- and 95-percent confidence limits are 0.872 and 0.844, so the inclusion of system results improved the assessment considerably.

Nonidentical Components -- Consider now a series system comprised of 2 nonidentical components, and suppose we have the following results:

	No. of Tests	No. of Successes
Component 1	20	19
Component 2	30	29
System	50	45

For these results, the normal equations are

$$\frac{64}{\hat{p}_1} - \frac{1}{1 - \hat{p}_1} - \frac{5\hat{p}_2}{1 - \hat{p}_1\hat{p}_2} = 0 ,$$

$$\frac{74}{\hat{p}_2} - \frac{1}{1 - \hat{p}_2} - \frac{5p_1}{1 - \hat{p}_1\hat{p}_2} = 0 .$$

The solution, obtained by the Newton-Raphson method, is $\hat{p}_1 = 0.9408$, $\hat{p}_2 = 0.9628$, and the estimated variance-covariance matrix is

$$\hat{V} = \begin{pmatrix} 0.00143976 & -0.00056347 \\ -0.00056347 & 0.00095783 \end{pmatrix} .$$

From this,

$$\hat{n}_1 = \frac{\hat{p}_1(1 - \hat{p}_1)}{0.00143976} = 38.68 , \quad \hat{x}_1 = \hat{p}_1\hat{n}_1 = 36.39 ,$$

and

$$\hat{n}_2 = \frac{\hat{p}_2(1 - \hat{p}_2)}{0.00095783} = 37.39 , \quad \hat{x}_2 = \hat{p}_2\hat{n}_2 = 36.00 .$$

From these, 90- and 95-percent lower MML confidence limits on p_1 are 0.859 and 0.836 and for p_2 are 0.886 and 0.864. The corresponding MMLI limits are 0.869 and 0.847, 0.866 and 0.843. Note that the inclusion of the system tests increased the information on Component 1 more than on Component 2. This is because Component 1 had the lower estimated reliability and therefore was more likely to have caused the system failures.

References

1. Buehler, R. J., "Confidence Limits for the Product of Two Binomial Parameters," Journal of the American Statistical Association, 52, pp. 482-493, 1957.

2. Easterling, R. G., "Approximate Confidence Limits for System Reliability," submitted for publication, 1970.

3. Easterling, R. G., and Prairie, R. R., "Combining Component and System Information," to appear in Technometrics.

4. Lipow, M., and Riley, J., Tables of Upper Confidence Limits on Failure Probability of 1, 2, and 3 Component Serial Systems, Vol. 1 and 2, Space Technology Laboratories, AD-609-100, AD-636-718, 1960.

5. Madansky, A., "Approximate Confidence Limits for the Reliability of Series and Parallel Systems," Technometrics, 7, pp. 495-503, 1965.

6. Myhre, J. M., and Saunders, S. C., "Comparison of Two Methods of Obtaining Approximate Confidence Intervals for System Reliability," Technometrics, 10, pp. 37-49, 1968.

7. Rosenblatt, Joan R., "Confidence Limits for the Reliability of Complex Systems," Statistical Theory of Reliability, ed. by Marvin Zelen, University of Wisconsin Press, 1963.

8. Steck, G. P., "Upper Confidence Limits for the Failure Probability of Complex Networks," SC-4133(TR), Sandia Laboratories, 1957.

DISTRIBUTION:
J. M. Wiesen, 100
W. C. Myre, 1210
R. L. Peurifoy, Jr., 1220
D. M. Olson, 1510
T. B. Lane, 1540
W. A. Gardner, Actg., 1550
H. E. Lenander, 1600
I. M. Moore, 1610
B. E. Arthur, 1640
H. E. Walker, 1641
R. E. Butler, 1642
R. R. Prairie, 1643
H. E. Anderson, 1643
R. G. Easterling, 1643 (50)
I. J. Hall, 1643
D. D. Sheldon, 1643
C. R. Clark, 1643
W. E. Boyes, 1644
W. L. Stevens, 1650
T. M. Burford, 1700
C. Winter, 1710
D. R. Morrison, 1720
R. G. Clem, 1730
H. D. Sivinski, 1740
A. A. Lieber, 1750
C. M. McCampbell, 2310
K. L. Gillespie, 2320
J. P. Shoup, 2330
M. T. Abegg, 2340
W. C. Kraft, 2440
T. S. Church, 2450
G. H. Roth, 2480
L. J. Paddison, 2490
C. M. Tapp, 2610
G. W. Rodgers, 2620
E. G. Franzak, 2630
J. A. Hood, 2650
L. J. Heilman, 7400
J. R. Meikle, 7410
R. H. Schultz, 7420
W. J. Denison, 7421
D. E. Murphy, 7422
R. D. Christopher, 7423
F. W. Muller, 7425
L. A. Bruckner, 7425
J. R. Ellesfon, 7425
J. N. Deverman, 7425
J. R. Sublett, 7430
D. E. Gregson, 8130
O. H. Schreiber, 8139
J. C. King, 8300
J. W. Weihe, 8320
M. L. Kramm, 9210
J. C. Eckhart, 9220
H. H. Patterson, 9230
R. E. Hepplewhite, 9420
G. C. McDonald, 3416 (3)
R. S. Gillespie, 3411
W. K. Cox, 3428-2 (15)