NASr-21(11)

MEMORANDUM
RM-4317-1-NASA
OCTOBER 1965

N 65-13518

/19

RELIABILITY GROWTH DURING A DEVELOPMENT TESTING PROGRAM

Richard E. Barlow and Ernest M. Scheuer

GPO PRICE $ ________

CFSTI PRICE(S) $ ________

PREPARED FOR:

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

The RAND Corporation
SANTA MONICA • CALIFORNIA

NASr-21(11)

MEMORANDUM
RM-4317-1-NASA
OCTOBER 1965

RELIABILITY GROWTH DURING A DEVELOPMENT TESTING PROGRAM

Richard E. Barlow and Ernest M. Scheuer

This research is sponsored by the National Aeronautics and Space Administration under Contract No. NASr-21. This report does not necessarily represent the views of the National Aeronautics and Space Administration.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

PREFACE

The original version of this RAND Memorandum, RM-4317-NASA, derived statistical methods for estimating reliability under growth assumptions which merely require that changes made to the system being tested do not make it worse. This revised version of the Memorandum differs from the original, which it supersedes, in having an improved conservative lower confidence bound for system reliability (Sec. V herein and in the original), and in eliminating Sec. VI of the original, which contained some peripheral and erroneous material.

While this Memorandum is addressed primarily to statisticians, it should also be of interest to test engineers and managers concerned with assessing a system's reliability. The investigation was undertaken as a part of the Apollo Contingency Planning Study, which RAND conducted for Headquarters, NASA, under Contract NASr-21(09); the additional work included in this revised version was done under Contract NASr-21(11).

One of the authors, Richard E. Barlow, is a consultant to The RAND Corporation.

SUMMARY

This study examines the problem of estimating the reliability of a system that is undergoing development testing. In such a program, changes are made to the system from time to time in order to increase its reliability. This study assumes that these changes are at least not deleterious, but, unlike some previous work in this area, it does not assume that system modifications cause reliability growth according to a prescribed functional form. The method described herein does, however, require that each failure be classified either as inherent or as reflecting a correctable cause.

The study proceeds on the supposition that the test program is conducted in K stages, with similar items being tested within each stage. For each stage, the number of inherent failures, of assignable cause failures, and of successes is recorded. It is supposed that the probability of an inherent failure, q_o, remains the same throughout the test program and that the probability of an assignable cause failure in the i-th stage, q_i, does not increase from stage to stage of testing. This Memorandum obtains maximum likelihood estimates of q_o and of the q_i's subject to the condition that they be non-increasing, and also obtains a conservative lower confidence bound for r_K, the reliability of the system in its final configuration of the test program. Numerical examples to illustrate these methods are given.

CONTENTS

I. INTRODUCTION

It is common practice, during the development of a system, to make engineering changes as the program develops. These changes are generally made in order to correct design deficiencies and, thereby, to increase reliability. This elimination of design weaknesses is what we mean by reliability growth.

The concept of reliability growth has been discussed by several authors. We mention some studies with which we are familiar.

Lloyd and Lipow in Chapter 11 of their book [1] give a model in which a system has only one failure mode; if the system operates successfully at a trial, no redesign action is taken prior to the next trial. If it fails at a trial, the designers attempt a modification which has a given probability of being successful. This model leads to an exponential growth model of the form.

$$R_n = 1 - A\,e^{-C(n-1)} , \tag{1}$$

in which R_n is the reliability of the system at the n-th trial and A and C are parameters to be estimated. Lloyd and Lipow also consider a situation in which a test program is conducted in N stages, each stage consisting of a certain number of trials of the item under test. All tests in a given stage of testing involve similar items. The results of each stage of testing are used to improve the item for further testing in the next stage. They impose on the data a reliability growth function of the form

$$R_k = R_\infty - \alpha/k , \tag{2}$$

where R_k is the reliability during the k-th stage of testing, and R_∞ (the "ultimate" reliability) and α are parameters. Lloyd and Lipow give maximum likelihood and least squares estimates of R_∞ and α and a lower confidence limit for R_k. Finally, they suggest some other forms that reliability growth models might take.

Wolman [2] considers a model in which a distinction is made between inherent (random) failures and assignable cause failures. He supposes that the number of design weaknesses (the source of assignable cause failures) is known and that they all have the same probability of causing a failure on a particular trial. Further, once a design weakness is observed, it is eliminated and will never again cause a system failure. Wolman is interested in calculating quantities such as the probability of eliminating all design weaknesses in n trials (assuming as known the probabilities of the two kinds of failure) -- not in estimating the parameters of his model. That is, his is a probabilistic model. (A statistical model is discussed below.)

Madansky and Peisakoff [3] have examined some data from Thor and Atlas Missile flights. They, too, distinguished between inherent and assignable cause failures and batched together data from comparable test units, but made no explicit use of any statistical model.

H. K. Weiss [4] has considered reliability growth as a process by which the mean time to failure of a system with exponential failure distribution is increased by removing failure causes during a development program.

Two related papers which have appeared recently are by Bresenham [5] and Corcoran, Weingarten, and Zehna [6].

II. A TRINOMIAL MODEL FOR RELIABILITY GROWTH

We propose the following model for a development program experiencing reliability growth. The test program is conducted in K stages. At each stage of experimentation, tests are run on similar items. The results of each stage of testing are used to improve the item for further testing in the next stage. We record for the i-th stage the number, a_i, of inherent failures,* the number, b_i, of assignable cause failures,** and the number, c_i, of successes. The probability of an inherent failure, q_o, is assumed to be constant and not to change from stage to stage of testing. The probability of an assignable cause failure in the i-th stage is q_i. Each trial results in exactly one of the outcomes: success, inherent failure, or assignable cause failure. We assume that the sequence of the q_i's is non-increasing. This means that changes made between stages of testing are not harmful to the system. The probability of success or the reliability in the i-th stage is, of course, $r_i = 1 - q_o - q_i$. By "reliability growth" we mean that the r_i's increase from stage to stage. This is accomplished by a decrease in the q_i's which must be brought about by appropriate engineering modifications of the system. We will obtain maximum likelihood estimates of q_o, and the q_i's under the restriction that they are non-increasing, and a conservative lower confidence interval for r_K, the reliability of the system in its final configuration of the test program.

---

*Failures that reflect the state-of-the-art and whose elimination would require an advancement thereof.

**Those which can be corrected by equipment or operational modifications.

It is worth remarking that the number of stages, K, and the number of trials, $a_i + b_i + c_i$, in the i-th stage may be fixed in advance or they may be random variables. Whichever the case, it will not alter the likelihood function corresponding to the experimental outcome on which our estimation procedure is based.

Let us compare our model with some of the others mentioned in the "Introduction." It shares with the work of Weiss, Madansky and Peisakoff, and Wolman the property that two types of failures (inherent and assignable cause) are distinguished.* Unlike Wolman, we do not suppose in our model that the number of assignable cause failures is known in advance or that each has the same probability of causing a failure. Like Lloyd and Lipow (in their model leading to our Eq. (2)) and Madansky and Peisakoff, we consider that test data are batched according to stages of sampling of homogeneous test items. Unlike Lloyd and Lipow, (i.e., Eq. (2)) we do not impose an arbitrary growth pattern on our test results.

---

*We will demonstrate the importance of this feature later by constructing a situation where, without this distinction, a nonsensical result obtains.

III. THE LIKELIHOOD FUNCTION AND THE MAXIMUM LIKELIHOOD ESTIMATES

The likelihood function corresponding to a_i inherent failures, b_i assignable cause failures, and c_i successes in stage i, i = 1, ..., K is

$$L(a_1,b_1,c_1,\ldots,a_K,b_K,c_K;\ q_o,q_1,\ldots,q_K) =$$

(3)
$$\prod_{i=1}^{K} \frac{(a_i + b_i + c_i)!}{a_i!b_i!c_i!}\ q_o^{a_i}\ q_i^{b_i}\ (1 - q_o - q_i)^{c_i}\ .$$

Upon differentiating the log likelihood with respect to q_o and q_i and setting the derivatives equal to zero, we find for the maximum likelihood estimates

(4)
$$\hat{q}_o = \sum_1^K a_i \bigg/ \sum_1^K (a_i + b_i + c_i) \quad ,$$

and

(5)
$$\hat{q}_i = (1 - \hat{q}_o)\ b_i \big/ (b_i + c_i) \ , \quad i = 1,\ldots,K \ .$$

Equations (5) are the maximum likelihood estimates of the q_i's in general. We want to obtain maximum likelihood estimates of the q_i's subject to the condition $q_1 \geq q_2 \geq \ldots \geq q_K$. This corresponds to our assumption that reliability does not decrease from stage to stage of testing. Adapting a result of Ayer, <u>et</u> <u>al</u>. [7] will give us these. Let $\bar{q}_1, \bar{q}_2, \ldots, \bar{q}_K$ denote the maximum likelihood estimates of $q_1, q_2, \ldots, q_K$ subject to the condition $q_1 \geq q_2 \geq \ldots \geq q_K$. If $\hat{q}_1 \geq \hat{q}_2 \geq \ldots \geq \hat{q}_K$, then $\bar{q}_i = \hat{q}_i$, $i = 1, \ldots, K$. If $\hat{q}_j < \hat{q}_{j+1}$ for some j ($j=1, \ldots, K-1$), then combine the observations in the j-th and (j + 1)-st stages and compute the maximum likelihood estimates of the q_i's by Eq. (5) for the K-1 stages thus formed. This procedure is continued until the estimates of the q_i's form a non-increasing sequence. These estimates are the maximum likelihood estimates of the q_i's subject to $q_1 \geq q_2 \geq \ldots \geq q_K$. An explicit expression can also be given for $\bar{q}_i$, i.e.,

$$\bar{q}_i = \max_{s \geq i} \min_{r \leq i} \frac{b_r + \ldots + b_s}{b_r + c_r + \ldots + b_s + c_s} \cdot (1 - \hat{q}_o)$$

(See [7]). We will illustrate this procedure in Sec. IV, but first we justify its validity.

Fix q_o and rewrite the likelihood Eq. (3) as

$$(6) \quad L = \left[q_o^{\Sigma a_i} (1-q_o)^{\Sigma(b_i+c_i)} \prod \frac{(a_i+b_i+c_i)!}{a_i! b_i! c_i!} \right] \prod \left(\frac{q_i}{1-q_o} \right)^{b_i} \left(1 - \frac{q_i}{1-q_o} \right)^{c_i} .$$

Letting $p_i = q_i/(1-q_o)$ so that $p_1 \geq p_2 \geq \ldots \geq p_K$, noting that $p_i \in [0,1]$ since $q_i \in [0, 1-q_o]$, and that the maximization of L with respect to the q_i does not involve the term in square brackets in Eq. (6), we find that we are in precisely the situation discussed by Ayer, <u>et</u> <u>al</u>. [7]. They find the maximum likelihood estimates for the p_i's subject to the constraint $p_1 \geq p_2 \geq \ldots \geq p_K$. Thus, their maximum likelihood estimate of p_i is $(1-q_o)$ times our maximum likelihood estimate of q_i. Maximizing on q_o we obtain the maximum likelihood estimate, Eq. (4), for q_o and are led to the maximum likelihood estimates for the q_i's given in the preceding paragraph.

The maximum likelihood estimate for system reliability at the i-th stage is

$$\hat{r}_i = 1 - \hat{q}_i - \hat{q}_o .$$

Note that assignable cause failures may "mask" inherent failures and vice versa. For example, if all failures in every stage were assignable cause failures, then $\hat{q}_o = 0$. This need not mean, however, that the inherent reliability is very high since elimination of assignable cause failures could subsequently lead to a high percentage of inherent failures. For this reason $\hat{r}_i$ is perhaps the only estimate to be "trusted."

IV. AN EXAMPLE

Suppose that a development testing program yielded the results shown in Table 1.

Table 1

Stage i	Inherent Failures a_i	Assignable Cause Failures b_i	Successes c_i	Trials $a_i + b_i + c_i$	$\frac{b_i}{b_i + c_i}$
1	0	1	0	1	1
2	0	1	0	1	1
3	0	1	0	1	1
4	1	1	1	3	1/2
5	0	1	4	5	1/5
6	0	1	0	1	1
7	0	1	0	1	1
8	0	1	3	4	1/4
9	9	1	27	37	1/28
Totals	10	9	35	54	--

Each stage of sampling, except the last, was terminated when an assignable cause failure occurred. A re-design effort was undertaken to eliminate the cause of failure, so that the test units in the succeeding stage were different from the earlier units but homogeneous in any given stage. We remark that this is the defining property of a stage, namely the homogeneity of all test units therein.

Note first that $\hat{q}_o = 10/54 = .1852$. To construct the maximum likelihood estimates for the q_i's subject to the condition that they be non-increasing, we must combine stages where there is a reversal of nonincreasingness of the ratios $b_i/(b_i + c_i)$ until we get a non-increasing sequence. It suffices to look at these ratios since the estimate of $\hat{q}_o$ does not depend on the grouping of the data into stages. Table 2 indicates how this grouping is done.

Table 2

i	b_i	c_i	$\frac{b_i}{(b_i + c_i)}$	First Combination	Second Combination
1	1	0	1	1	1
2	1	0	1	1	1
3	1	0	1	1	1
4	1	1	1/2	1/2	1/2
5	1	4	1/5 }	1/3 }	3/7
6	1	0	1 }		
7	1	0	1	1 }	
8	1	3	1/4	1/4	1/4
9	1	27	1/28	1/28	1/28

Observe that

$$\frac{b_5}{b_5 + c_5} < \frac{b_6}{b_6 + c_6} \quad ,$$

so that we combined stages 5 and 6. There is yet a reversal between the ratios for the new fifth stage and the new sixth stage (the original seventh stage) so we next combine those stages. We now have eliminated all reversals and obtain as maximum likelihood estimates $\bar{q}_i$ of the q_i's subject to $q_1 \geq q_2 \geq \ldots \geq q_K$,

$$\bar{q}_1 = \bar{q}_2 = \bar{q}_3 = 22/27 = .8148,$$
$$\bar{q}_4 = 11/27 = .4074,$$
$$\bar{q}_5 = \bar{q}_6 = \bar{q}_7 = 22/63 = .3492,$$
$$\bar{q}_8 = 11/54 = .2037, \text{ and}$$
$$\bar{q}_9 = 11/378 = .0291 .$$

Thus the maximum likelihood estimate for r_9, the reliability of the system in its final test configuration, is

$$\bar{r}_9 = 1 - \hat{q}_o - \bar{q}_9 = .7857 .$$

If no assumption of reliability growth were made -- that is, if all test units were (incorrectly) supposed to be homogeneous and if no distinction were made between inherent and assignable cause failures -- the estimate of reliability would be

$$\hat{r}_9 = \sum c_i \Big/ \sum (a_i + b_i + c_i) = 35/54 = .6481 .$$

V. A CONSERVATIVE LOWER CONFIDENCE BOUND FOR SYSTEM RELIABILITY

In this section we do not need to distinguish between inherent and assignable cause failures. We consider, as before, a model in which the reliability (probability of success) does not decrease from stage to stage of testing; that is, $r_1 \le r_2 \le \ldots \le r_K$. We seek a procedure for obtaining a $100(1-\alpha)$ per cent lower confidence bound on r_K, the reliability of the system in its final configuration of the test program.

We will show that the procedure is to treat the data as though they were homogeneous, that is, as if no reliability growth were taking place, and use the standard technique to obtain a one-sided lower confidence limit on a binomial parameter having observed s successes in n trials.

Explicitly, having observed s successes in n trials, a $100(1-\alpha)$ per cent one-sided lower confidence limit, p_L, for the binomial parameter p is the solution of the equation

$$\sum_{j=s}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \alpha \quad , \tag{7}$$

Equivalently p_L is the solution of

$$\sum_{j=o}^{s-1} \binom{n}{j} p^j (1-p)^{n-j} = 1 - \alpha \tag{8}$$

if $s \ge 1$. If $s = 0$, p_L is customarily taken to be zero for each α.

Returning to the problem at hand, if c_i successes are observed in n_i trials in stage i $(i = 1, \ldots, K)$, and writing

$$S = \sum_{1}^{K} c_i \quad \text{and} \quad n = \sum_{1}^{K} n_i \quad ,$$

then if $u \ge 1$, simple considerations of stochastic comparability yield

(9) $$P[S \le u - 1 | r_1 \le \ldots \le r_K = r_*] \ge P[S \le u - 1 | r_1 = \ldots = r_K = r_*] =$$

$$\sum_{j=o}^{u-1} \binom{n}{j} r_*^j (1 - r_*)^{n-j} .$$

Define u(r) as the smallest u satisfying

(10) $$\sum_{j=o}^{u-1} \binom{n}{j} r^j (1 - r)^{n-j} \ge 1 - \alpha ,$$

and define r^o as the largest r such that

(11) $$u(r) - 1 = S .$$

Now we have

(12) $$P[S \le u(r_K) - 1 | r_1 \le \ldots \le r_K] = P[r_K \ge r^o | r_1 \le \ldots \le r_K]$$

since u(r) increases with r. Using equations (9) and (12) and the definition of u(r), we conclude

(13) $$P[r_K \ge r^o | r_1 \le \ldots \le r_K] \ge 1 - \alpha .$$

Thus, to obtain a 100(1-α) per cent lower confidence bound on r_K having observed c_i successes in n_i trials in stage i (i = 1, ..., K), we set $S = \sum_1^K c_i$, $n = \sum_1^K n_i$ and find r_o, the largest r such that

(14) $$\sum_{j=o}^{S-1} \binom{n}{j} r^j (1 - r)^{n-j} \ge 1 - \alpha$$

and then claim that r_o is a $100(1-\alpha)$ per cent lower confidence bound on r_K in accordance with Equation (13).

This lower confidence bound on r_K is the best that can be achieved under our assumptions, since equality will be attained in inequality (9) when all the r_i's are equal; that is, when there is no reliability growth present.

Example: In the development testing program cited in Sec. IV, 35 successes were observed in 54 trials in nine stages of testing. Using binomial tables, we find that $r^o = .53$ at the 95 per cent confidence level. That is, we are 95 per cent confident that $r_9 \geq .53$. This is a conservative statement as we noted earlier. It is, in fact, the same estimate we could obtain assuming no reliability growth; i.e., $r_1 = r_2 = \ldots r_K$.

Note that if one looks at only the results of stage 9, standard methods yield, for 27 successes in 37 trials, a lower 95% confidence limit of .58 for the reliability in stage 9. This merely shows that if enough data are available from the last stage, the standard binomial approach may be preferred. Our method, however, enables one to use the data from all stages.

VI. BINOMIAL VERSUS TRINOMIAL MODELS

In this section we construct an example to show that it would not suffice to consider a binomial model for reliability growth using the maximum likelihood approach of Sec. III. We do this by exhibiting an experimental outcome in which the maximum likelihood estimate for r_n under a binomial model is nonsensical while the corresponding estimate under our trinomial model is eminently reasonable.

Suppose we consider a binomial model in which we make no distinction between inherent and assignable cause failures. Denote the probability of success at trial i by r_i and assume as before that $r_1 \le r_2 \le \ldots \le r_n$. The unrestricted maximum likelihood estimate of r_i is 0 or 1, according as failure or success is observed at trial i. To obtain the maximum likelihood estimates of the r_i's under the restriction $r_1 \le r_2 \le \ldots \le r_n$, one invokes the procedures of Ayer, *et al*. [7], which we used in Sec. III. However, if the n-th trial results in a success, the maximum likelihood estimate of r_n will be unity -- independent of what transpired on earlier trials. In particular, this would be the maximum likelihood estimate of r_n even if all trials prior to the n-th had been failures. On the other hand, if the n-th trial were a success, our trinomial model would give as the maximum likelihood estimate for r_n the proportion of successes observed in the n trials when all recorded failures are inherent failures.

VII. A TREND TEST FOR RELIABILITY GROWTH

In the foregoing we have assumed that the probability of an assignable cause failure does not increase during the development testing program. We feel that the validity of this hypothesis would be determined on the basis of engineering knowledge. However, we propose a test for reliability growth; that is, for non-increasingness of the q_i's.

Mann [8] has given two tests against downward trend and provided tables for their use. Specifically, we are given data $X_1, X_2, \ldots, X_n$ in that order. The null hypothesis is that the X's are randomly arranged. The alternative hypothesis is that X_i has continuous cumulative distribution function F_i, with $F_i(t) < F_{i+k}(t)$ for every i, every t, and every $k > 0$; that is, the sequence of the X_i's is stochastically decreasing. To test against upward trend, merely test $-X_1, -X_2, \ldots, -X_n$ against downward trend.

Suppose each stage of testing is terminated when an assignable cause failure occurs. We identify X_i with the number of trials since the last assignable cause failure. The X_i's should increase if reliability growth is taking place. Note, however, that here we are dealing with a discrete random variable so that the c.d.f. will not be continuous. We can circumvent this problem by adding a uniform [0,1] random variable to each of the random variables suggested above without changing the appropriate probabilities under the null hypothesis. We can then apply one of the Mann procedures to this new random variable.

REFERENCES

1. Lloyd, David K., and Myron Lipow, Reliability: Management, Methods, and Mathematics, Prentice-Hall, Englewood Cliffs, 1962.

2. Wolman, William, "Problems in System Reliability Analysis," in Marvin Zelen (ed.), Statistical Theory of Reliability, University of Wisconsin Press, Madison, 1963.

3. Madansky, Albert, and M. P. Peisakoff, Thor and Atlas In-flight Reliability Growth(U), The RAND Corporation, RM-2598, June, 1960, (Confidential).

4. Weiss, Herbert K., "Estimation of Reliability Growth in a Complex System with a Poisson-Type Failure," Operations Research, Vol. 4 (1956), pp. 532-545.

5. Bresenham, Jack E., Reliability Growth Models, Technical Report No. 74, Department of Statistics, Stanford University, August, 1964.

6. Corcoran, W. J., H. Weingarten, and P. W. Zehna, "Estimating Reliability After Corrective Action," Management Science, Vol. 10 (1964), pp. 786-795.

7. Ayer, Miriam, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman, "An Empirical Distribution Function for Sampling with Incomplete Information," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 641-647.

8. Mann, H. B., "Nonparametric Tests Against Trend," Econometrica, Vol. 13 (1945), pp. 245-259.