1964-07

Estimating Reliability After Corrective Action

Corcoran, W.J.; Weingarten, H.; Zehna, P.W.

Corcoran, W. J., H. Weingarten, and P. W. Zehna. "Estimating reliability after corrective action." Management Science 10.4 (1964): 786-795.
http://hdl.handle.net/10945/59829

MANAGEMENT SCIENCE
Vol. 10, No. 4, July, 1964
Printed in U.S.A.

ESTIMATING RELIABILITY AFTER CORRECTIVE ACTION*†

W. J. CORCORAN[1], H. WEINGARTEN[2], AND P. W. ZEHNA[3]

A model is defined wherein corrective action may be accounted for in improving the estimation of reliability over the usual nominal success ratio. Probabilities for correcting any one of K failure modes which may arise are assumed known and a multinomial sampling procedure is discussed. Mean reliability is defined as a function of the unknown probabilities attached to the failure modes and the problem of estimating this parameter is posed.

Two procedures which have been typically used in the past are discussed and some obvious weaknesses are pointed out. We then proceed to investigate several new estimators which we propose for the problem. No attempt is made to optimize with regard to a choice of estimators but properties of each of them are discussed which point to their limitations as well as usefulness in a given problem.

Finally, we discuss some of the implications of the model we have assumed and propose other models that might be worthy of merit. Recommendations are made for further research in this timely and important area of investigation.

1. Formulation of a Model

The problem which we wish to consider in this paper is one which may arise in the final stage of development of an expensive item. Suppose that such an item has been developed to the point where it is necessary to observe its performance and establish its reliability. Consequently, the item will be subjected to testing and we consider that on the basis of N tests we may observe a total of N_0 successes yielding a nominal reliability of N_0/N. However, the causes of some or all of the $N - N_0$ possible failures could be identified and corrective action could be taken to attempt to eliminate these failure modes. Now we would like to "take credit" for the improvement in reliability which should be achieved by such corrective action.

One perfectly straightforward way of estimating current reliability after the N tests have been performed is to simply conduct additional tests after the corrective action has been taken. If N' additional tests are performed and N'_0 successes are observed, then the ratio N'_0/N' may be used to estimate current reliability. Alternatively, since the reliability may be treated as a binomial parameter, standard confidence interval estimates may be used and we are on

* Received May 1963.

† This work was sponsored by the U.S. Navy under Contract No. NOW 61-0153-c (FMB).

[1] Navy Special Projects Office, presently with the United Technology Corporation, Sunnyvale, California.

[2] Navy Special Projects Office, presently with the Center for Naval Analyses, Washington, D. C.

[3] C-E-I-R, Inc., Los Altos, California, presently with the U.S. Naval Postgraduate School, Monterey, California.

firm ground, statistically speaking. However, the cost of the items to be tested may be such as to make this direct approach prohibitively expensive, or possibly there is a shortage of the items and operational considerations may preclude their use for testing purposes at this time. Consequently, we pose the following problem: Assuming we have confidence in our knowledge of the effectiveness of the contemplated corrective action which may be taken on observed failure modes, how should we use the results of the first N tests to draw inferences about current reliability?

To cast the problem within the framework of statistical estimation, we make the following set of assumptions. A given test may result either in success with unknown probability p_0 (the initial reliability) or exactly one of K fixed but unspecified failure modes where the unknown probability of a failure of type i is denoted q_i and $p_0 + \sum_{i=1}^K q_i = 1$. The N tests to be performed shall be independent and so our underlying probability model is multinomial with parameters $N, p_0, q_1, q_2, \cdots, q_K$. Accordingly, we denote the number of observed successes in the N tests by N_0 while N_i is used to denote the number of observed failures of type i so that $N_0 + \sum_{i=1}^K N_i = N$.

By fixing K, we tacitly assume that no new failure modes are introduced by corrective action. Also, no corrective action is to be taken until all N tests are performed so that our sampling procedure is based on a fixed sample size N rather than being sequential. We do assume, however, that when a failure occurs it can be classified as to type. With regard to corrective action, we assume that if a failure of type i is observed, then there is a *known* probability a_i of removing that mode of failure by corrective action. Consequently $a_1, a_2, \cdots a_K$ are conditional probabilities, conditioned on observing the various modes of failure and must be based on some information external to our present model in order to be treated as known quantities.

Given that N tests are to be performed and attempts made thereafter to correct those failure modes that may be observed, we define the following random variable as a measure of the current reliability.

$$p_0^* = p_0 + \sum_{i=1}^{K} y_i q_i, \quad \text{where} \quad y_i = \begin{cases} 0 & \text{if } N_i = 0 \\ a_i & \text{if } N_i > 0 \end{cases} \tag{1.1}$$

Intuitively we see that p_0^* measures current reliability by adding to the initial reliability a proportionate amount of the failure probabilities corresponding to the observed failure modes after corrective attempts on those modes have been made. Since $E[y_i] = a_i[1 - (1 - q_i)^N]$ for $i = 1, 2, 3, \cdots, K$, the expected value of p_0^*, called mean reliability and denoted μ, is given by,

$$\mu = p_0 + \sum_{i=1}^K a_i q_i [1 - (1 - q_i)^N] \tag{1.2}$$

The above set of assumptions and definitions outlines one of several possible models and we wish to consider now the problem of estimating mean reliability on the basis of the observed random variables $N_0, N_1, N_2, \cdots, N_K$ within the structure of this model. It should be clear that we are undertaking to develop a course of action before the tests are performed and before corrective action is taken also, so that we have in hand in advance a procedure we plan to follow after the testing and corrective action are over. All future discussion should be considered in this context.

2. The Estimation Problem

Having defined our model and established an estimation problem, let us observe at the outset that the relative sizes of N and K will play a significant role in the behavior of the estimators we treat. For example, if $K = 1{,}000$ and $N = 30$ then we are essentially faced with the problem of estimating 1,000 (distinct) parameters with a sample size of 30. Obviously no estimator can be expected to be particularly sharp in such circumstances.

Strictly speaking, we do not estimate p_0^* since it is a random variable and not a parameter. Consequently, our estimation procedures will be based upon estimating the parameter μ as defined in (1.2). On the other hand, the variance of p_0^* can be calculated in a straightforward manner and we find that it is given by,

$$\begin{aligned}\sigma^2[p_0^*] &= \textstyle\sum_{i=1}^K a_i^2 q_i^2 (1 - q_i)^N [1 - (1 - q_i)^N] \\ &\quad + 2 \textstyle\sum_{i<j} a_i a_j q_i q_j [(1 - q_i - q_j)^N - (1 - q_i - q_j + q_i q_j)^N]\end{aligned} \tag{2.1}$$

Moreover, with the q's and the a's fixed, $\sigma^2[p_0^*] \to 0$ as $N \to \infty$. Consequently, for large values of N, the distribution of p_0^* is concentrated at μ and so an estimate of μ also provides a prediction for the actual value of p_0^* in this case. Asymptotically then, we may speak somewhat loosely of estimating p_0^*.

At least two approaches to our problem have been used in the past. In the first approach, the attitude is that failure modes which may be observed, which we plan to correct, should be converted to success since a test on which a failure would occur, would not be a success insofar as that particular failure mode is concerned. However, since the probability of correcting a given failure mode is usually different from unity, an adjustment is made by averaging out according to the a's. This is essentially how p_0^* is defined but of course we cannot compute the value of p_0^* after experimentation because of the unknown parameters. Using estimates of the parameters in (1.1) the estimator used in this approach is defined by,

$$p_1 = N_0/N + \textstyle\sum_{i=1}^K y_i N_i/N \tag{2.2}$$

Since $y_i N_i \equiv a_i N_i$, we may also write,

$$p_1 = N_0/N + \textstyle\sum_{i=1}^K a_i N_i/N \tag{2.3}$$

Using the fact that the marginal distribution of N_i is binomial, expectations are easily computed and the bias of p_1, $b(p_1)$, as an estimate of μ is found to be,

$$b(p_1) = \textstyle\sum_{i=1}^K a_i q_i (1 - q_i)^N \tag{2.4}$$

Since $b(p_1) > 0$ we see that p_1 tends to overestimate μ although the bias decreases toward zero with increasing sample size N. However, examples can be chosen in which the amount of overestimation is a serious matter and certainly detracts from the usefulness of p_1 as an estimator. In particular, suppose $a_i = 1$ and N is small relative to K. Then p_1 takes on the value unity regardless of the outcome of the experiment. But clearly the amount of improvement in reliability cannot be too great since only a relatively small number of failure modes can be observed in the first place.

A second approach to the problem is based on the idea that, since the cause of a given failure will be eliminated by corrective action, the test on which that failure may occur will now be considered as "no-test" and hence removed from consideration in estimating the current situation. Again, averaging with respect to the a's to account for the uncertainty in corrective action, the following estimator is defined so as to incorporate these ideas.

$$p_2 = N_0/(N - \textstyle\sum_{i=1}^{K} a_i N_i) \tag{2.5}$$

Intuitively we see that p_2 accounts for the improvement due to corrective action by dividing the number of successful tests by a number smaller than the total number of tests thus giving a larger value than would be obtained using the nominal reliability N_0/N. Moreover, it is a simple matter to verify that $p_2 \leqq p_1$ so that p_2 is more conservative than p_1 and would thus tend to overcome the overestimation inherent in p_1.

Nevertheless, p_2 exhibits a behavior similar to p_1 in the case $a_i \equiv 1$, for in that case $p_2 = 1$ whenever $N_0 > 0$ while it is undefined when $N_0 = 0$. As another (perhaps pathological) example which brings out the unsatisfactory nature of p_2, consider the case where $p_0 = 0$ and hence, appropriately, $N_0 = 0$. Further, let $q_1 = q_2 = \cdots = q_{10} = .1$, $a_1 = a_2 = \cdots = a_{10} = .99$ and $N = 100$. Now there will be 100 failures distributed among $h \leqq 10$ failure modes and, regardless of that distribution, $\sum_{i=1}^{10} a_i N_i = 99$. In this example, $\mu = \sum_{i=1}^{10} (.99)(1 - (.9)^{100}) \doteq .99$, while $p_2 = 0/(100 - 99) = 0$. This is certainly not a satisfactory result. We might observe that in this example $p_1 = .99$ for all outcomes of the testing procedure.

Because of these and other examples which tend to discredit the use of either p_1 or p_2, we began a search for new estimators which would overcome some of these undesirable properties. The results of our combined efforts are summarized in the next section.

3. Some Estimators and Their Properties

The first estimator we treat is defined from the maximum likelihood point of view. To this end, we replace the parameters in the expression for μ by their respective maximum likelihood estimators to obtain,

$$p_3 = N_0/N + \textstyle\sum_{i=1}^{K} a_i N_i/N[1 - (1 - N_i/N)^N]. \tag{3.1}$$

We observe that p_3 is an improvement over p_1 and p_2 with regard to the overestimation discussed in the last section, particularly in the case $a_i \equiv 1$. However, direct computation of moments for p_3 to find exact expressions were found to be intractable. If we assume that N is sufficiently large so that $(1 - N_i/N)^N$ is approximately $\exp[-N_i]$, then we find, using the binomial marginal distribution of N_i,

(3.2) $$E(p_3) \approx p_0 + \textstyle\sum_{i=1}^{K} a_i q_i - e^{-1} \sum_{i=1}^{K} a_i q_i [1 - q_i(1 - e^{-1})]^{N-1}$$

Equation (3.2) promptly led us to modify p_3 slightly by a factor to compensate for e^{-1} in the above expression and we define,

(3.3) $$p_4 = N_0/N + \textstyle\sum_{i=1}^{K} a_i N_i/N - (N/(N-1))^N \sum_{i=1}^{K} a_i N_i/N(1 - N_i/N)^N$$

Accordingly, we may write,

(3.4) $$E(p_4) \approx p_0 + \textstyle\sum_{i=1}^{K} a_i q_i - \sum_{i=1}^{K} a_i q_i [1 - q_i(1 - e^{-1})]^{N-1}$$

The expression given in (3.4) is at most equal to μ so that p_4 tends to underestimate and thus provides for a certain degree of conservatism. These results are, however, only asymptotic and the required magnitude of N may be prohibitively large, particularly in the kind of model that we have described.

Turning to unbiasedness as a criterion, we use the expansion

$$(1 - q_i)^N = \sum_{j=0}^{N} (-1)^j \binom{N}{j} q_i^j$$

to write μ in the form,

(3.5) $$\mu = p_0 - \sum_{i=1}^{K} \sum_{j=1}^{N} (-1)^j \binom{N}{j} a_i q_i^{j+1}$$

Hence, finding an unbiased estimator for μ resolves itself into finding an unbiased estimator for q_i^{j+1}, $j = 1, 2, 3, \cdots, N$; $i = 1, 2, 3, \cdots, K$. But for each i, N_i has a binomial distribution and, using the factorial moment generating function for that distribution, it is well known that

$$E\left[\binom{N_i}{r}\right] = \binom{N}{r} q_i^r$$

for $r = 1, 2, 3, \cdots, N$. Moreover, it is a simple matter to prove that an unbiased estimator for q_i^{N+1} does not exist when the sample size is N. This can be seen as follows. Consider a binomial situation with N and p. Let $f(x)$ be the estimator of p^{N+1}. We must then have identically

$$\sum_{x=0}^{N} f(x) \binom{N}{x} p^x q^{N-x} \equiv p^{N+1}$$

However, the quantity on the left is a polynomial in p of degree at most N. Thus no $f(x)$ exists. The extension to the multinomial case is straightforward. Consequently we cannot find an unbiased estimator for μ using the random variables $\binom{N_i}{r}$. Nevertheless, using the fact that $0 \leqq N_i \leqq N$ and $\binom{N_i}{r} = 0$ whenever $N_i < r$, we define an estimator by,

(3.6) $$p_5 = N_0/N - \sum_{i=1}^{K} \sum_{j=1}^{N-1} (-1)^j (j+1)/(N-j) a_i \binom{N_i}{j+1}$$

Using the facts we mentioned above, it follows that,

$$E(p_5) = p_0 - \sum_{i=1}^{K} \sum_{j=1}^{N-1} (-1)^j \binom{N}{j} a_i q_i^{j+1} \tag{3.7}$$

It then follows by subtraction that the bias of p_5 is given by,

$$b(p_5) = (-1)^N \textstyle\sum_{i=1}^{K} a_i q_i^{N+1} \tag{3.8}$$

Thus we see that p_5 tends to overestimate if N is even and underestimate if N is odd. Also, $b(p_5) \to 0$ as $N \to \infty$ so that p_5 is asymptotically unbiased. Even for moderate values of N, however, it is clear that the amount of bias may be insignificant from a practical point of view. Notably, significant bias would occur when the a's are all very nearly unity and the q's are all relatively large (K being relatively small in that case). For many examples we have examined (for instance $K = 10$, $q_i = .09$ and $p_0 = .1$) we find that the bias is zero to 3 decimal places even for $N = 5$.

With regard to the problem of overestimation, we observe that $j + 1$ is at least 2 in the expression on the right in (3.6). Hence the term $\binom{N_i}{j+1} = 0$ for all j whenever N_i is zero or one and so a given N_i must be at least 2 before any contribution to a change in the estimate of mean reliability over the nominal reliability is allowed. In this sense, p_5 is conservative. On the other hand, it is easy to construct examples in which p_5 may assume a value greater than unity. Primarily this will occur when the a's are all very nearly one and the q's are relatively large. However, in such a case we may from a practical point of view simply call our estimate one. Such truncation will affect the bias slightly but will also reduce the variance.

Regarding variance and other moments, we find the calculations to be somewhat intractable. Writing

$$\frac{1}{N-j} = \int_0^1 x^{N-j-1}\,dx$$

and using familiar combinatorial formulas, p_5 may be written,

$$p_5 = N_0/N + \sum_{i=1}^{K} a_i N_i/N + \sum_{i=1}^{K} (-1)^{N_i} y_i \binom{N}{N_i}^{-1} \tag{3.9}$$

where y_i is defined as in Section 1.[4] This form of p_5 was found to be more suitable for computational purposes, particularly for computer simulations which were carried out at NWL Dahlgren, Va. and C-E-I-R, Los Angeles.

A third estimator which we have examined for this problem was devised to incorporate some of the desirable properties which p_5 possesses and to give a simpler form at the same time. In particular we were guided by the fact that p_5 gives no credit to a failure mode which occurs but once (a property which consistently appeared in those estimators which we considered to be otherwise de-

[4] This reduction was pointed out to us by Dr. Mel Peisakoff of C-E-I-R, Los Angeles.

sirable). Accordingly, we define,

$$p_6 = N_0/N + \sum_{i=1}^{K} z_i N_i/N \quad \text{where} \quad z_i \begin{cases} a_i & \text{if } N_i > 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

Again, using the marginal distribution of N_i, it is easily verified that the mean of p_6 is given by,

$$E[p_6] = p_0 + \sum_{i=1}^{K} a_i q_i - \sum_{i=1}^{K} a_i q_i (1 - q_i)^{N-1} \tag{3.11}$$

A simple calculation then yields the bias of p_6.

$$b(p_6) = -\sum_{i=1}^{K} a_i q_i^2 (1 - q_i)^{N-1} \tag{3.12}$$

Now $b(p_6) < 0$ for all N and $b(p_6) \to 0$ as $N \to \infty$ so that p_6 tends to underestimate μ and is asymptotically unbiased.

Because of the relatively simple form of p_6 it is also possible to obtain exact expressions for the second moment. The calculations are laborious but perfectly straightforward, using only the joint trinomial distribution of the pair (N_i, N_j) when $i \neq j$. We find that,

$$E[p_6^2] = (p_0 + \sum_{i=1}^{K} a_i q_i)^2 + f(N)/N \tag{3.13}$$

where $f(N) = O(1)$. Since $E[p_6] \to p_0 + \sum_{i=1}^{K} a_i q_i$ as $N \to \infty$, it follows that $\sigma^2[p_6] \to 0$ as $N \to \infty$ where $\sigma^2[p_6]$ is the variance of p_6. Since $b(p_6) \to 0$ also, we may say that p_6 is a consistent estimate of μ, i.e., $p_6 \xrightarrow{p} \mu$.

As a matter of further interest, we have established bounds on the bias of p_6 which enable us to assess the behavior of p_6 in extreme cases. Observing that $| b(p_6) | \leqq \sum_{i=1}^{K} q_i^2 (1 - q_i)^{N-1}$, a simple application of the method of Lagrange reveals that the expression on the right side of the inequality, as a function of the q's with the linear constraint $\sum_{i=1}^{K} q_i = 1 - p_0$, is maximized when $p_i \equiv (1 - p_0)/K$. Consequently,

$$| b(p_6) | \leqq \frac{(1 - p_0)^2}{K} \left(1 - \frac{1 - p_0}{K}\right)^{N-1} \leqq \frac{1 - p_0}{N} \left(1 - \frac{1}{N}\right)^{N-1}. \tag{3.14}$$

Letting $B = (1 - 1/N)^{N-1}/N$, the following table demonstrates the manner in which B varies with the sample size.

TABLE 3.1

N	B
2	.25
3	.15
4	.11
5	.082
10	.039
25	.015
50	.0075

Thus we see, for example, that when N is as large as 25 and p_0 is as large as $\frac{1}{3}$, p_6 will tend in the long run to underestimate μ by no more than .01 even in the "worst" case of the q's.

The final estimator which we present was defined from a more conservative point of view. Thus, the loss which results from using an estimator which gives an estimate that is too high may be relatively more important than the gain achieved by the increase over merely using the nominal reliability N_0/N. The latter procedure represents the ultra-conservative point of view since corrective action is ignored in that case. Accordingly, we define,

$$p_7 = N_0/N + \sum_{i=1}^{K} y_i(N_i - 1)/N \tag{3.15}$$

where, as before,

$$y_i = \begin{cases} a_i & \text{if} \quad N_i > 0 \\ 0 & \text{if} \quad N_i = 0 \end{cases}.$$

We observe that p_7 also ignores failure modes that occur but once and a simple calculation reveals,

$$E[p_7] = p_0 + \sum_{i=1}^{K} a_i q_i - (1/N) \sum_{i=1}^{K} a_i[1 - (1 - q_i)^N] \tag{3.16}$$

By subtraction we find that

$$b(p_7) = \sum_{i=1}^{K} a_i q_i (1 - q_i)^N - (1/N) \sum_{i=1}^{K} a_i[1 - (1 - q_i)^N] \tag{3.17}$$

It is easily verified that,

$$p_6 - p_7 = \sum_{i=1}^{K} z_i/N \quad \text{where} \quad z_i = \begin{cases} a_i & \text{if} \quad N_i > 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.18}$$

Consequently, $E(p_7) \leqq E(p_6)$ so that $b(p_7) \leqq b(p_6)$. Hence $b(p_7) < 0$ for all N, is more conservative than p_6, and $b(p_7) \to 0$ as $N \to \infty$.

As before, with details omitted, we find

$$E[p_7^2] = (p_0 + \sum_{i=1}^{K} a_i q_i)^2 + g(N)/N \tag{3.19}$$

where $g(N) = 0(1)$. Putting this fact together with (3.15) and (3.16), $\sigma^2[p_7] \to 0$ so that p_7 is also a consistent estimate of μ.

4. Concluding Remarks

The estimators presented in the preceding section are by no means the only ones considered. Most of the others, however, contain such obvious defects in the light of the model we established that they were rejected as unsuitable. While we make no claim for completeness by any means, we do feel that the estimators given in Section 3 provide a good approach to the problem of point estimation that we have outlined. Moreover, we feel that we have provided some degree of versatility with regard to various criteria that are usually adopted for point estimation, viz., unbiasedness, consistency, conservatism, maximum likelihood. Judgments as to the choice of the estimator we have presented must then depend upon the user's choice of criteria. For reasons peculiar to Navy

Special Projects, the estimator p_6 is presently being used in actual applications of the problem presented here.[5]

Perhaps the most incomplete feature of our work is the lack of distribution theory. Efforts to the contrary, we have not been able to establish any significant progress in the direction of probability distributions for any of the estimators we have treated. This of course seriously hampers the problem of confidence interval estimation. One approach would be to assume normality and thereby obtain normal confidence intervals for μ. The results, even if validated, would nevertheless be asymptotic and we feel that one should exercise extreme caution in the application of asymptotic results within the model we have described. Alternatively, standard non-parametric methods might be applied to the estimators to arrive at confidence intervals.

With regard to our model, we feel that we must stress the obvious and remark that our estimators are no better than the probabilities $\{a_i\}$ of successful corrective action which are assumed. In the real world it is often difficult to obtain an entirely satisfactory diagnosis of failure. Corrective action is sometimes subject to performance penalties, and a compromise action may be chosen, leaving some doubt as to its effectiveness. Occasionally, steps taken to eliminate one failure mode lead to the introduction of another. Furthermore, the test of a device such as we have been discussing often proceeds in phases and a failure in an early phase may preclude entry of that failure mode in a later phase. When this is the case, our model would be more appropriate as a means of estimating mean reliability for a particular phase. Overall reliability is then a function (such as the product in the case of independence) of the phase reliabilities. All of these factors tend to limit the model we have assumed. Nevertheless, we feel that, within the sphere of conventional analysis, a good case can be made for the model.

Another model which might be examined has also occurred to us. Rather than view the problem as parametric estimation, one might define the true reliability within the same general structure of the present model. Such a reliability will of course be a random variable and will be dependent on whether or not a given failure mode is actually removed. Then, viewing the problem as one of preduction rather than estimation, various predictors might be judged against some loss function such as squared error. Of course the estimators treated here are also candidates as predictors in such a problem.

Finally, another model worthy of investigation would be one in which corrective action is taken sequentially. Thus, instead of waiting for N tests to be performed as we have insisted upon in our model, one might assume that when a single test results in failure, an immediate diagnosis takes place followed by the corresponding corrective action. As tests are performed and corrective action thus attempted the problem of both parametric estimation and prediction might

[5] Further details on the use of p_6 may be found in a Navy Special Projects Office report entitled, "Estimation of Reliability after Corrective Action on Observed Failure Modes," June 1962.

be examined. Also, cost of sampling might be "built in" to such a model. If any a priori information about the q's is available one might also examine the problem from the Bayes approach.

We wish to thank the referee for his extremely helpful comments on the original version of this paper.

References

1. MOOD, A. M., *Introduction to the Theory of Statistics*, McGraw-Hill Book Company, Inc., New York, 1950.
2. PARZEN, E., *Modern Probability Theory and Its Applications*, John Wiley and Sons, Inc., New York, 1960.
3. WILKS, S. S., *Mathematical Statistics*, John Wiley and Sons, Inc., New York, 1962.