

AD-A092 658

FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS

F/6 12/1

OPTIMAL SAMPLE DEPENDS ON OPTIMALITY CRITERION.(U)

OCT 80 H LACAYO, C A PEREIRA, F PROSCHAN

AFOSR-78-3678

UNCLASSIFIED

FSU-STATISTICS-M-559

AFOSR-TR-80-1205

NL

[]

ALL INFORMATION CONTAINED HEREIN IS UNCLASSIFIED



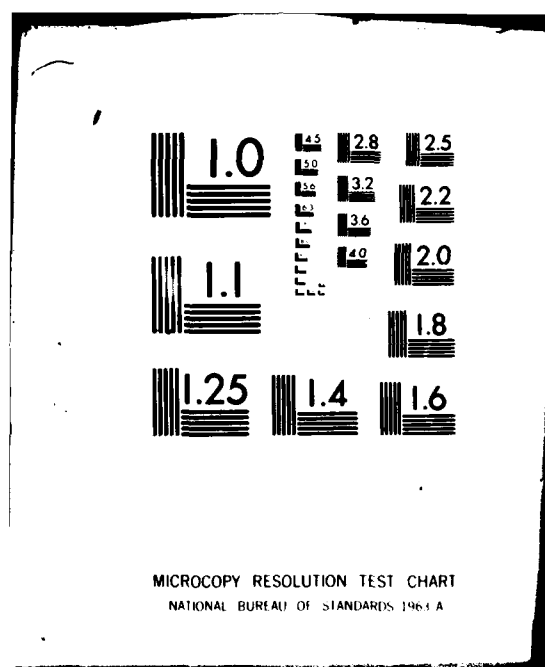
END

DATE

FILMED

1-8

DTIC



SECURITY CLASSIFICATION OF THIS PAGE **UNCLASSIFIED**

LEVEL II

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER AFOSR-TR-80-1205	2. GOVT ACCESSION NO. AD-A092658	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) OPTIMAL SAMPLE DEPENDS ON OPTIMALITY CRITERION.		5. TYPE OF REPORT & PERIOD COVERED Interim	
6. AUTHOR(s) Herbert/Lacayo, Carlos/Pereira, Frank/Proschan, and Carl Samdal		7. PERFORMING ORG. REPORT NUMBER	
8. CONTRACT OR GRANT NUMBER(s) AFOSR-78-3678		9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
10. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University Department of Statistics Tallahassee, Florida 32306		11. REPORT DATE October 1980	
12. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, Washington, DC 20332		13. NUMBER OF PAGES 6	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED	
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE			

6. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

**DTIC
ELECTE
DEC 8 1980
S D B**

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Sampling, Optimal, sequential testing, Bayesian statistics
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Four different sampling procedures are compared. Each is optimal under an appropriate optimality criterion.

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A092658

DDC FILE COPY

Optimal Sample Depends on Optimality Criterion

by
Herbert Lacayo¹, Carlos A.B. Pereira², Frank Proschan¹, and
Carl Särndal

Statistics Report M -- 559
AFOSR Technical Report No. 78--116

October, 1980
Department of Statistics
The Florida State University

¹Research performed under the support of the Air Force Office of Scientific Research, AFSC, USAF, Grant [REDACTED]. *AFOSR 78-3678*

²Research sponsored by CAPES and USP- Brasil.

³University of Montreal and Florida State University.

Key Words: Sampling, optimal, sequential testing, Bayesian statistics.

AMS 1970 Subject Classifications: 62D05, 62N05.

Approved for public release
distribution unlimited.

80 12 08 108

optimal procedure is described in one of the first three statements above. The Bayesian procedure proposed in statement 4 is appropriate for λ unknown.

Let us take the easy cases first. Suppose our goal is to maximize the number of batteries that we know are functioning at time t_2 . Then it is easy to see that we should follow Procedure 2. The probability that a battery, found to be functioning at time t_1 , will be found upon inspection at time t_2 to be functioning is $e^{-\lambda t_2}/e^{-\lambda t_1} = e^{-\lambda(t_2 - t_1)}$. On the other hand, a previously uninspected battery has probability $e^{-\lambda t_2}$ of being in the functioning state upon inspection at time t_2 . Since $e^{-\lambda t_2} < e^{-\lambda(t_2 - t_1)}$, it follows that the optimal strategy to maximize the number of batteries we know to be functioning at time t_2 is as described in Procedure 2.

Suppose, on the other hand, our goal is to weed out as many defective batteries as possible by inspection. Now, the optimal plan is described in Procedure 1. This follows from the fact that the probability that a battery found to be functioning at time t_1 , will be found upon inspection at time t_2 to be in the failed state is $(e^{-\lambda t_1} - e^{-\lambda t_2})/e^{-\lambda t_1} = 1 - e^{-\lambda(t_2 - t_1)}$. On the other hand, a previously uninspected battery has probability $1 - e^{-\lambda t_2}$ of being in the failed state upon inspection at time t_2 . Since $1 - e^{-\lambda t_2} > 1 - e^{-\lambda(t_2 - t_1)}$, it follows that the optimal strategy to weed out the maximum number of defective batteries is as described in Procedure 1.

A third possible goal is to estimate as precisely as possible - precision is defined by the squared error loss function - the number X of functioning batteries at time t_2 when λ is known. The problem is to select a sample

of fixed size, from among (a) the batteries already inspected at time t_1 and found to be still alive and (b) the batteries not yet inspected at time t_1 .

Note that the (Bernoulli) variable used to indicate the state (functioning or not) of a battery selected from among (a) has variance $e^{-\lambda(t_2 - t_1)}(1 - e^{-\lambda(t_2 - t_1)})$, while the variable used to indicate the state of a battery selected from among (b) has variance $e^{-\lambda t_2}(1 - e^{-\lambda t_2})$. Since we wish to obtain an estimate of the number X with smallest possible risk, we simply follow procedure 3 above.

4' Suppose that λ is unknown and we choose a distribution to represent our prior opinion about λ . In particular, suppose we choose a Gamma distribution a priori. That is,

$$\pi(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\lambda} \quad (1)$$

is the prior density for λ . Having observed a sample of size n_1 at time t_1 containing f_1 failed batteries and $n_1 - f_1$ functioning batteries, we compute the posterior mean of the function:

$$p_1(1 - p) - p_2(1 - p_2) = e^{-\lambda t_2}(1 - e^{-\lambda t_2}) - e^{-\lambda(t_2 - t_1)}(1 - e^{-\lambda(t_2 - t_1)}) = e^{-\lambda t_2} - e^{-2\lambda t_1} - e^{-\lambda(t_2 - t_1)} + e^{-2\lambda(t_2 - t_1)} \quad (2)$$

where p_1 and p_2 are defined in statement 3 above. If the result obtained is greater than zero, we select one battery from among the $N - n_1$ completely untested batteries and test it at time t_2 ; if the result obtained is less than or equal to zero, we test at t_2 a battery chosen from among the $n_1 - f_1$ batteries that survived the test at time t_1 . We incorporate this new result in the original sample and obtain an adjusted likelihood. With

this likelihood and the original prior we again compute the posterior mean of (2) in order to decide from which set (a) or (b) the next battery will be selected. This procedure will be followed sequentially to complete the sample of size n_2 or until either sets (a) or (b) are exhausted; in this case complete the sample (of size n_2) with items chosen from the remainder set (either (a) or (b)).

To specify the computations explicitly we list the following notation:

- n_1 = number of batteries tested at time t_1 .
- f_1 = number of batteries (among these n_1) failed at time t_1 .
- n_{12} = number of batteries tested at time t_1 and at time t_2 .
- f_{12} = number of batteries failed at time t_2 from among these n_{12} .
- n_2 = number of batteries tested at time t_2 .
- f_2 = number of failures among the $n_2 - n_{12}$ batteries tested only at time t_2 . (Note that the f_{12} batteries are not part of the f_2 batteries.)

For reasons of simplification we relax the above notation in the following formulas. The quantities n_2 , n_{12} , f_2 , and f_{12} are viewed below as changing steadily at each step of the sequential procedure.

The general likelihood is given by:

$$L = (e^{-\lambda t_1})^{n_1 - n_{12} - f_1} (1 - e^{-\lambda t_1})^{f_1} (e^{-\lambda t_2})^{n_2 - f_2 - f_{12}} (1 - e^{-\lambda t_2})^{f_2} \times (e^{-\lambda t_1} - e^{-\lambda t_2})^{f_{12}} \quad (3)$$

We emphasize the fact that this likelihood is computed at each step with the current values of n_2 , n_{12} , f_2 , and f_{12} .

From the likelihood (3) and the prior (1), we may compute the Bayes estimator for $e^{-\lambda(c_2 t_2 - c_1 t_1)}$, as follows:

$$E\{e^{-\lambda(c_2 t_2 - c_1 t_1)} | \text{Data}\} =$$

$$= \frac{1}{A} \sum_{i=0}^{f_1} \sum_{j=0}^{f_2} \sum_{k=0}^{f_{12}} \frac{(-1)^{i+j+k} \begin{pmatrix} f_1 \\ i \end{pmatrix} \begin{pmatrix} f_2 \\ j \end{pmatrix} \begin{pmatrix} f_{12} \\ k \end{pmatrix}}{[\beta + t_1(n_1 - f_1 + i + k - c_1) + t_2(n_2 - f_2 + f_{12} + j - k + c_2)]^\alpha}, \quad (4)$$

$$\text{where } A = \sum_i \sum_j \sum_k \frac{(-1)^{i+j+k} \begin{pmatrix} f_1 \\ i \end{pmatrix} \begin{pmatrix} f_2 \\ j \end{pmatrix} \begin{pmatrix} f_{12} \\ k \end{pmatrix}}{[\beta + t_1(n_1 - f_1 + i + k) + t_2(n_2 - f_2 + f_{12} + j - k)]^\alpha}.$$

Note that the final Bayes estimator for $e^{-\lambda t_2}$ is computed by taking $c_2 = 1$ and $c_1 = 0$ in (4).

We emphasize that the Bayes sequential procedure described above may be only an approximation to the optimal Bayes procedure. We do not know if this procedure, although very intuitive, is the one which minimizes the risk for the squared loss function.

Final Remark:

Note that the criteria used depend only on the values of p_1 and p_2 , not on the fact that we have an exponential distribution. This shows the generality of the methods discussed. On the other hand, suppose that q_1 , q_2 , and $q_3 = 1 - (q_1 + q_2)$ are the probabilities of a battery failing respectively, before t_1 , in the interval (t_1, t_2) , and after t_2 . If we represent our prior opinion about (q_1, q_2, q_3) by a Dirichlet distribution, then the Bayes solution in this case is very similar to the one presented here.