	REPORT DOCU	MENTATION	PAGE		
a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE	MARKINGS		
a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION	AVAILABILITY	OF REPORT	
		DISTRIBUT	ION UNLIMI	TED .	
B. DECLASSIFICATION / DOWNGRADING SCHEDU					
. PERFORMING ORGANIZATION REPORT NUMBE	ER(S)	5. MONITORING	ORGANIZATION	REPORT NUMBE	R(S)
a. NAME OF PERFORMING ORGANIZATION Deputy Under Secretary of the Army (Operations Research)	6b. OFFICE SYMBOL (If applicable) SAUS-OR	7a. NAME OF M	ONITORING ORG	JANIZATION	ECTE 198
A ADDRESS (City, State, and ZiP Code)		75. ADDRESS (C/	ty, State, and Zi	(P Code)	10100
HODA ATTN: SAUS-OR, RM 2E660 Washington, D.C. 20310-0102				S	Nº 8E
a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (if applicable)	9. PROCUREMEN	TINSTRUMENT	IDENTIFICATION-1	NUMBER
L ADDRESS (City, State, and ZIP Code)	.*	10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT
Persining II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel Ba. TYPE OF REPORT Final Final SUPPLEMENTARY NOTATION	e Reduced by Sec OVERED TO	quential Anal 14. DATE OF REPO 30 Septembe	ysis DRT (Year, Mont) er 1984	h, Day) 15. PAG	E COUNT
Persining II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel Ba. TYPE OF REPORT Final S. SUPPLEMENTARY NOTATION COSATI CODES FIELD GROUP SUB-GROUP	e Reduced by Sec OVERED TO AB. SUBJECT TERMS (Reliability	quential Anal 14. DATE OF REPO 30 Septembe Continue on reverse	ysis PRT (Year, Month er 1984 e if necessary a > Bay	h, Day) 15. PAG nd identify by bl yesian -8tat	E COUNT ock number) istics ;
Persining II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel Ba. TYPE OF REPORT Final S. SUPPLEMENTARY NOTATION COSATI CODES FIELD GROUP ABSTRACT (Continue on reverse if necessary)	e Reduced by Sec OVERED TO TO Reliability Sequential Ana Truncated Aest and identify by block	quential Anal 14. DATE OF REPO 30 September Continue on reverse alysis; t) -	PRT (Year, Month or 1984 e if necessary and Bay Fig	h, Day) 15. PAG nd identify by bl yesian Stat sher's Exac 2 DC	E COUNT ock number) istics ; t Pest ;
Pershing II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel a. TYPE OF REPORT Final S. SUPPLEMENTARY NOTATION COSATI CODES FIELD ABSTRACT (Continue on reverse if necessary) To replace the erroneous us bissile system in annual tests, two efficient methods, both tre attention is called to the Trun essentially the same results, a given level of confidence) and advantages of the Bayesian meth- sampling and other uses of seque	e Reduced by Sec OVERED TO TO TO TO TO TO TO TO TO TO	Quential Anal 14. DATE OF REPO 30 Septembe Continue on reverse alysis; t)- number) xact Test to eloped, in co tist and Baye l Probability reduction in th out. They a	e if necessary a derive the plaboration ratio Tesion the size of are general	h, Day) 15. PAG nd identify by bl yesian Stat sher's Exac 2 DC	E COUNT ock number) istics; t Pest; y of a isticians; ular thods gave s (for a rogram. Th le for lot F(1,2,2)
Pershing II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel a. TYPE OF REPORT Final Supplementary NOTATION COSATI CODES FIELD ABSTRACT (Continue on reverse if necessary) To replace the erroneous us hissile system in annual tests, wo efficient methods, both tre attention is called to the Trun essentially the same results, a given level of confidence) and advantages of the Bayesian meth- sampling and other uses of seque	e Reduced by Sec OVERED TO TO TO TO TO TO TO TO TO TO	Quential Anal 14. DATE OF REPO 30 Septembe Continue on reverse alysis; t)	e if necessary and derive the pllaboration r Ratio Tesion the size of are general	nd identify by bly yesian Stat sher's Exac 2 0 2	E COUNT ock number) istics; t Pest; y of a isticians; ular thods gave s (for a rogram. Th le for lot F(1,2)/
Persining II Folow-On Test: Siz Personal AUTHOR(S) Willard, Daniel Ba. TYPE OF REPORT Final S. SUPPLEMENTARY NOTATION COSATI CODES FIELD GROUP SUB-GROUP COSATI CODES FIELD ABSTRACT (Continue on reverse if necessary To replace the erroneous us hissile system in annual tests, wo efficient methods, both tre attention is called to the Trun essentially the same results, a piven level of confidence) and advantages of the Bayesian meth bampling and other uses of seque D. DISTRIBUTION/AVAILABILITY OF ABSTRACT	e Reduced by Sec OVERED TO TO TO TO TO TO TO AND Sequential And Truncated Arest and identify by block e of Fisher's E: the author deve ated by frequent cated Sequential nd so justified there with a rec ods are pointed ential tests.	21. ABSTRACT SE UNCLASSIF	e if necessary and ber 1984 e if necessary and balaboration esian methon r Ratio Tesion the size of are general costs of are general	nd identify by bly yesian Stat sher's Exac $\angle D \subseteq - \leq$ reliabilit n with stat ds. Partic t. Both me of the test the test p ly applicab	E COUNT ock number) istics ; t Pest, y of a isticians, ular thods gave s (for a rogram. Th le for lot F(1),2); ;

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

PERSHING II FOLLOW-ON TEST:

SIZE REDUCED BY SEQUENTIAL ANALYSIS



by



Daniel Willard

September 1984

Chapter I	The Problem
Chapter II	Mathematical Preliminaries
Chapter III	Launer and Singpurwalla's Proposal
Chapter IV	Woodroofe's Proposal
Chapter V	Other Stopping Criteria
Chapter VI	Conclusion
Appendices	A References B Bibliography C Extract from WSEG 92C D HP-41 Programs E Supporting Chapter III

۴ ┥

SUBJECT: Pershing II Follow-On Test: Size Reduced by Sequential Analysis

By memorandum of 30 August 1982 (Reference 1), the Under Secretary of the Army tasked the service to "review our [operational test] methodology, to include considerations of mathematical rigor, risks, planning horizon, costs, and operational matters." In discussion of this matter with the author, he further elaborated the objectives:

a) Minimize cost of testing over the program life. Monitor all test results, including those of components as well as of the system, to minimize "no-tests" and to save on full-up tests. Use sequential analysis to further pare requirements for missile flights.

b) Criteria of test adequacy should be no more severe than those of other services (e.g., Minuteman, Poseidon).

c) Challenge the necessity for an annual update.

d) Consider whether testing, maintenance float, and reload were independent requirements as opposed to multiple missions for the same inventory of missiles.

The task was passed to the Army Research Office (Research Triangle, NC) which manages the business of the Army's Mathematics Steering Committee (Dr. Jagdish Chandra, Chairman), supporting mathematical research of relevance to the Army and the improvements in mathematical methods employed in the Army's research and study agencies.

The work summarized here is composed of contributions of several statisticians whose aid was solicited by the AMSC: Dr. Michael Woodroofe (University of Michigan)*, Dr. Nozer Singpurwalla (George Washington University), and Dr. Robert Launer (Army Research Office), as well as the author of this report. Others have provided informal comments and criticisms. An early version of this paper, prior to the author's knowledge of this other research, was presented as a talk at a conference of Army mathematicians (Reference 2).

* At Rutgers University during the course of this research.

Chapter I

The Problem

Two documents combined set forth the guidance the Joint Chiefs of Staff have provided to the military services regarding the conduct and reporting of tests of certain systems. For the Army only the Pershing Missile system is covered (Pershing I and Ia, and now Pershing II).

In a memorandum of 1975 (Reference 3), the Joint Chief of Staff directed that numerical confidence statements should be based on WSEG Report 92C (Reference 4), an extract of which is at Appendix C. "The goal of a test program should be to allow detection of a minimum change of X percent at the Y percent confidence level." * It suggests, by way of example, the use of Fisher's Exact Test to demonstrate success or failure in meeting this criterion.

References 3 and 4 have just been superseded. The revisions (References 5 and 6) eliminate an ambiguity and add considerations not previously called for and not discussed here except to note that the criteria to be applied to Pershing II are now less demanding than those applied to strategic systems. Fisher's Exact Test is still countenanced.

This use of this criterion appeared to the author to lack a sound statistical justification, and attempts to patch it up were unsuccessful. Appeal to a number of practicing statisticians within and outside the Army supported my challenge to Fisher's Exact Test (FET) in its application to Pershing reliability tracking. No one was contesting the ability of the FET to provide estimates of the probability that two samples, which have yielded pass-fail data, come from the same parent population, though Kendall and Stuart (Reference 7), do condemn its use for small samples.

With such an error apparently arising from an application of the methods of the "frequency" school of statistics, the obvious alternative was to try the methods of the "Bayesian" school.

There are many expositions of methods based on the use of Bayes' Theorem, the most recent of which--"Bayesian Reliability Analysis" by Martz and Waller--(Reference 8) I shall quote at intervals. Among the works arguing for the adoption of Bayesian methods, the following are noteworthy:

* X and Y are classified numbers.

Raiffa and Schlaifer - Applied Statistical Decision Theory (Reference 9) with a very complete description of the method of conjugate prior distributions.

Jaynes E.T., "Prior Probabilities" (IEEE Transactions on System Science and Cybernetics, September 1968) (Reference 10). Deduction from the principles of maximum entropy and invariance under certain group transformations leads directly to the Beta distribution as conjugate prior to a Bernoulli process; indeed to

$$dP(p;n,s) = p^{s-1} (1-p)^{n-s-1} dp / B(s,n-s)$$
 1.1

where s is the number of successes in n trials observed as the basis for estimating p. This removes some of the "ad hoc" or "mathematically convenient" color of conjugate priors when relying on Raiffa and Schlaifer.

Martz and Waller perhaps epitomize the case best:

"There are several benefits in using Bayesian methods in reliability. First of all, it is important to recognize that all statistical inferential theories, whether sampling theory, Bayesian, likelihood, or otherwise, require some degree of subjectivity in their use. Sampling theory requires assumptions concerning such things as a sampling model, confidence coefficient, which estimator to use, and so on. For example, a sampling theory analysis proceeds as if it were believed a priori that the data were exactly [exponentially] distributed, that each observation had exactly the same mean life θ , and that each observation was distributed exactly independently of every other sample observation. The Bayesian method provides a satisfactory way of explicitly introducing and organizing assumptions regarding prior knowledge or ignorance. These assumptions lead via Bayes' theorem to posterior inferences, that is, inference obtained once the data have been incorporated into the analysis, about the reliability parameter(s) of interest. Bayes' theorem provides a simple, error-free truism for incorporating the prior information. The engineering judgment and prior knowledge are brought out into the open and are there for everyone to see instead of being quietly hidden. The engineer usually appreciates this opportunity to divulge such prior information in a formalized way."

The authors I commend are not, on philosophical matters, in complete agreement, and the authors (and critics) of the methods proposed in this paper have their differences, some of which become important as we proceed. Suffice it to say that the Bayesian approach requires a more careful statement of the problem, to include in particular the prior distribution function, costs and risks: matters which the frequentists collapse into the confidence limits \propto and β . If there is indeed a legitimate uncertainty in (the form of) the prior distribution, that uncertainty must surely propagate into an uncertainty in the predictions for the process. In some cases results can be shown to be insensitive to the prior, and thus a convergence of Bayesian and frequentist answers occurs; but lacking such invariance, the frequentists are hard pressed to prove they have solved the right problem.

Having said this, I must confess that for some purposes we shall employ the frequentist approach, primarily because a full Bayesian solution has not been worked out.

Section 1. Literal Interpretation of JCS Guidance:

". . . annual . . . detection of a minimum reliability change of X percent at the Y percent confidence level."

A "change" in something means that its previous value has been defined. It would appear that an evaluation of the results of the first year's Follow-on-Test (FOT) is to be compared to that of the Operational Test (the base-line)(OT), and the evaluations of subsequent FOTs are to be compared to the evaluations made a year ago. The tests being of something less than the full combat mode of the system, projection to combat capability is to be made; thus while test results are to be reported, they are to be interpreted as well. This interpretation is surely to be based on all prior knowledge of system performance; i.e., all prior testing as well as that most recently at hand, "weighted" (one might say) by expert judgment of the relevance of older tests and analysis.

In the case of Pershing II, we shall have an inventory of missiles produced over a period of time and expected to be in service for a longer period. From the point of view of homogeneity, the inventory may need to be divided into two or more blocks, based on the significance of any changes in the production process during the run. When they are subjected to (annual) test, missiles will be of different ages as well from different blocks; so serial number and age may influence reliability at the time of testing or use in combat. It is clear, then, that in treating of a "change" in reliability, we are dealing with an uncertain base. Options which are open to us include:

a) Computing a "best" estimate from the OT firings, and treating it as the exact value of the reliability at that time of all the inventory. . b) Computing as in (a), but associating an uncertainty (standard deviation) to it also, to describe the uncertain reference point.

In either case, the results of each subsequent (annual) test would be compared to this as standard.

c) Computing as in (b), but then modifying the estimates using the results of subsequent tests (more trials, more successes, more failures). There are extremes in this process which are to be avoided:

(i) This modification might consist of using only the previous year's results as indication of the remaining inventory.

(ii) This modification might consist of accumulating the results of all prior tests, without regard to the aging effect or block modifications.

Judgment is clearly needed. Limiting the criterion to the smallness of the latest annual change (with small samples in the two cases) could result in a dangerous accumulation of change over the system life. On the other hand, where no statistically significant change has been detected, it would be reasonable to add one year's results to the results of the whole prior test series of a homogeneous block in estimating the average value at, say, the average age of the tested articles. It is probably not possible to specify in advance the details of the critical results to be reported. What is more important is that analyses be conducted to discover what are the constant and what are the variable components of the system reliability. Finally, detection of a trend should make it possible to forecast when the results of that trend will no longer be tolerable, and so signal the degree of urgency with which management should act to correct the trend.

This brings us to the question of the frequency of d) reporting the results of testing and analysis. The current practice is an annual report which probably has its roots in adminstrative cycles. The technical problems which reporting communicates to management are probably of two sorts: long-term aging with gradual deterioration, ("one-hoss shay" syndrome) and catastrophic failures. The latter tend to announce their presence in consistent repetitions of particular failure modes, and so call for out-of-cycle action no matter what the standard interval between reports. The former, on the other hand, are evidence of problems only slowly exacerbating, and so allow a more leisurely pace of administrative response. Alternatives to the present annual cycle are proposed below, for situations in which no guarantee of a clear bright green light or red light is available annually: (i) A guarantee can be given of a low likelihood of having to wait more than, say, 16 months for such a signal, along with the provision of a technical review of all failures showing any repetitions of mode. (ii) Administratively, skipping one year's report may be simpler.

. These options will be explored in one or more places in the mathematical sections to follow.

Two assumptions have immediately to be disposed of:

1) Because Fisher's Exact Test is mentioned in JCS guidance, its use is correct and mandatory.

Fisher's Exact Test is an enumeration of all possible relative outcomes in two series of pass-fail tests, subject to the restraints that the numbers of tests in each series be fixed and the combined number of successes also. It yields the probability that the articles tested in the two series were drawn from the same population--one with a fixed probability of pass. If the total number of successes is not controlled, the results of FET admit of this interpretation only in the limit of large samples. Given that the probability of success could be different in the two populations, it is sometimes claimed that FET can be used to estimate the probability that they differ by prescribed amounts. This claim is unwarranted. The JCS could be faulted for suggesting the test, but they did not underwrite the extended use as in the Army's methodology. (See Kendall and Stuart; also Chapter III).

2) We can know the reliability of an object.

We shall never know the "true" as-manufactured reliability of the components of the Pershing system, and much of such knowledge as we do gain will come at the expense of tactical inventory. It may be that, for the purposes of designing tests of operational reliability, we need not know this a priori probability with any great accuracy; and so methods which treat it as known for this purpose may be satisfactory. This does not justify the assumption when analyzing the results of actual tests.

Section 2. Mathematical Preliminaries

Bayes' Theorem: The Need for a Prior Distribution

Essential to much of what follows is Bayes' Theorem, sketched here as background. The conditional probability of an event B, given that another event A has occurred, is symbolized and defined by

$$P(B|A) = \frac{P(A,B)}{P(A)} \qquad 1.2$$

where P(A) ($\neq 0$) is the marginal probability of event A, and P(A,B) is the probability of joint occurrence of A and B. One may also speak of P(A/B) = P(A,B)/P(B) with similar meanings and limits, leading to

$$P(BIA) P(A) = P(AIB) P(B) \qquad 1.3$$

Given that B can occur in n ways Bi $(i=1,2,\ldots,n)$ one of which always occurs with A, we may sum expressions like Eq. 1.3 for the entire set of events Bi

$$P(A) \sum_{i} P(Bi|A) = \sum_{i} P(A|B_{i})P(B_{i}) = P(A) \quad 1.4$$

as the multiplier of P(A) is equal to 1, having encompassed all possible pairings. If $P(A) \neq 0$, we have Bayes' Theorem:

$$P(B_i|A) = \frac{P(A|B_i) P(B_i)}{\sum_i P(A|B_i) P(B_i)}$$
 1.5

Suppose now that events Bi are logically (causally) prior to event A. Then P(Bi) is called the prior distribution of Bi, P(A/Bi) the likelihood of A, given Bi, P(A) the marginal distribution of A, and P(Bi/A) the posterior distribution of Bi. Bayes' Theorem, given in symbols by Eq.1.5, may then be stated in words:

Posterior Distribution = <u>F.ior Distribution X Likelihood (Function)</u> Marginal Distribution

(This argument holds for both discrete and continuous distributions of probability.)

Likelihood functions are a familiar staple of probability theory, being forecasts of the frequency of chance events A based ct presumptions about certain prior events or conditions (a die that is unbiased, the "normal" distribution of errors, half-life of a know: radioactive substance). Marginal distributions then are forecasts of the results of experiments. Bayes' Theorem tells us that inferences about the events Bi which lead to a marginal distribution cannot be derived from the likelihood function alone, but require knowledge of the prior distribution P(Bi) as well. In the context of our task, we need to know more than the results of a set of missile firings to infer the reliability of the missile.

Other requirements of a Bayesian analysis will be discussed as the issues arise.

<u>Section 3.</u> <u>Illustration of an Analysis in Accord with JCS</u> <u>Guidelines</u>

We assume that the missiles and associated ground equipment used in an annual test do come from a homogeneous population, and that the several tests within that year are statistically independent. We assume further that the reliability p is definable, and then may assert that were we to know p, the probability of s,' successes and f,' failures in n,' trials (nl' = sl' + fl') would be by Bernoulli's formula (a likelihood function):

$$\binom{n'_{1}}{s'_{1}}p^{s'_{1}(1-p)^{s'_{1}}}$$
 where $\binom{n}{s} \equiv \frac{n!}{s!f!}$

From component testing, comparison with similar systems, comparison with other products of the same manufacturer, engineering analysis, we should develop an estimate of p and a measure of our confidence in that estimate. Methods exist, e.g. that of Maximum Entropy (Reference 10), for constructing from this information a function with the properties of a probability distribution--a prior distribution. Constraints of reasonableness and mathematical convenience come into the selection process. With limited information at hand, there may be no unique solution. The analyst is free to try several priors and to observe the sensitivity of answers to such variations.

Given a likelihood function, there can generally be found a "conjugate" prior function (so-called because it marries mathematically to the likelihood function); properly a class of such functions, dependent on a limited number of parameters to distinguish members of the class. Conjugate to the Bernoulli's distribution is the Beta distribution, written

$$dP(s_{o}, f_{o}) = p^{s_{o}-1} (1-p)^{f_{o}-1} dp / B(s_{o}, f_{o})$$
1.6
where $\int_{p=0}^{1} dP(s_{o}, f_{o}) = 1$, $B(s_{o}, f_{o}) = \frac{\Gamma(s_{o})\Gamma(f_{o})}{\Gamma(s_{o}+f_{o})}$,
and $\Gamma(n) = (n-1)!$ for n an integer.

Different sets of the parameters s_0 and f_0 give rise to functions whose graphs are variously peaked at some locale within the limits of 0 to 1, are relatively flat, are J-shaped and strongly peaked at 0 or 1, or are even U-shaped and strongly peaked at both 0 and 1. It is a rich set of functions.

Taking the product of $dP(s_0, f_0)$ with the Bernoulli function, we get

$$\binom{n'_{i}}{s'_{i}} p^{s'_{i}+s_{o}-1} (i-p)^{f'_{i}+f_{o}-1} dp / B(s_{o},f_{o})$$
 1.7

which when integrated over the range of 0 to 1 gives

$$\binom{\eta_{1'}}{s_{1'}}B(s_{1},f_{1})/B(s_{2},f_{2})$$
 where $\begin{cases} s_{1}=s_{2}+s_{1}'\\ f_{1}=f_{2}+f_{1}' \end{cases}$

the marginal distribution of s_1 ' given $B(s_0, f_0)$ as prior. The ratio of Eqs. 1.6 and 1.7 gives the posterior distribution of p for s_1 ' and f_1 ' observed:

$$p^{s_{i-1}}(1-p)^{f_{i-1}}dp/B(s_{i},f_{i})$$
 1.8

explaining my notation and revealing the meaning of conjugation.

From a prior distribution $B(s_0, f_0)$, and a likelihood function for a test of a sample of size n_1 ', we have created a function which, as a posterior distribution from that experiment, is logically the prior when testing a second sample of size n_2 '. This process can be repeated ad libitum, making sample 1 refer to all prior information and sample 2 the latest test. Now the JCS asks to know the probability that the reliability of sample 2 (and by inference that of the population from which it was drawn) is less than a certain fraction k (o < k \leq 1)of the reliability estimate p of sample 1. If the evidentiary basis for this answer lies entirely in the test of n2' items, then we may assume instead a uniform prior distribution, drop the primes on n2', s2', and f2' and represent this probability by

$$P(kp_{1}) = \int_{c}^{kp_{1}} p_{2}^{s_{2}-1} (1-p_{2})^{f_{2}-1} dp_{2} / B(s_{2},f_{2})$$

which we then integrate over the distribution of pl to get the probability that p2 \leq kpl:

.

$$P(p_{2} \leq k_{P_{i}}) = \int_{c}^{t} p_{i}^{s_{i}-1} (1-p_{i})^{f_{i}-1} P(k_{P_{i}}) dP_{i} / B(s_{i},f_{i}) \quad 1.9$$

The probability that p2>kpl is just 1 minus this result.

As an aid to understanding the generality of this result, consider the case where $pl = rl \times r3$ and $p2=r2 \times r3$ where r3 is a reliability factor not subject to degradation but just as much subject to discovery as rl and r2. Within the framework of Betafunction priors, we might be led to the posterior distribution:

$$dP = Kr_{1}^{s_{1}-1} (1-r_{1})^{f_{1}-1} r_{2}^{s_{2}-1} (1-r_{3})^{f_{2}-1} r_{3}^{s_{3}-1} (1-r_{3})^{f_{3}-1} dr_{1} dr_{2} dr_{3} \qquad 1.10$$

where s3(f3) is the total number of observed successes (failures) of the subsystems described by r3. For any values of r3 and k between 0 and 1, $P(p2 \le kp1) = P(r2 \le kr1)$. When the latter function is given by integrating Eq.1.10 first over r3 from 0 to 1, it is clear that the result is the same as though r3 = 1 (i.e., it can be ignored). Thus using the criterion $p2 \le kp1$ we cam be freed of any concern about reliability factors common to p1 and p2. I would assert that this is a good reason to employ this criterion in preference to the one described next.

The JCS guidance has not always been interpreted as speaking to a proportional reduction in reliability; sometimes it has been interpreted as measuring a reduction of, say, 100d percentage points.

Instead of Eq. 1.9 we would then use

$$P(p-d) = \int_{0}^{p-d} \frac{p_{2}^{s_{2}-i}(1-p_{1})^{f_{1}-i}dp_{2}}{p_{2}^{s_{1}-i}(1-p_{2})^{f_{1}-i}\int_{0}^{p-d} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{1}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{1}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{1}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{1}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{1}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})\int_{0}^{t} \frac{p_{2}^{s_{2}-i}(1-p_{2})^{f_{2}-i}dp_{2}}{B(s_{2},f_{2})}dp_{2}}$$

(While we have strayed from the neatness of conjugate functions, by reason of the incomplete integrals, we still have a consistent method. Similar expressions will be found in Reference 8, p. 271.)

* Indeed, the latest revision of the JCS guidance (Reference 5) mandates this form of the criterion. • Eqs. 1.9 and 1.11 give mathematical meaning to the JCS guidance. If at the chosen confidence level it is deemed that there has been no significant change in the reliability between samples 1 and 2, then sample 2 should be merged with sample 1 in preparation for the next year's testing. Other criteria should be examined also (e.g., probability that there has been no significant departure from a nominal value), but that does not refute the translation into mathematics of the JCS guidelines.

At this point I note that much of the historical course of development of mathematics has been devoted to a search for solutions requiring a minimum of actual manipulation of numbers. The approximations used by statisticians are simply good examples of this. The ready availability today of powerful computers reduces the need to employ approximations which may be questionable in particular cases. Most of the calculations to be described here have been carried out on a programmable hand calculator (HP-41) or home computer (Apple, Commodore, etc.). Accordingly, the reader need not be concerned with an apparent intractibility of the formulas. They could be evaluated in the field by the troops of a Pershing fire unit.

There are two matters of concern: the prior distribution and limits to the size of Sample 1. I have already discussed problems with the prior distribution. One assertion made is that with increase in the size of the data base it can become misleadingly narrow, ignoring "unknown-unknowns." A different way of saying this is that tests performed sufficiently long ago may be irrelevant in describing the present state of the missile inventory; the meaning of this argument is that a larger annual test size is needed to compensate for stale data in Sample 1. The question of test size will be the subject of the following chapters. Of course, if there is no evidence of a change in reliability over the years, there is no reason to purge old data.

Section 4. Optimum Test Size

In order to determine the number of missiles which must be procured in the next few years to support a test program through a long period of service life, one must have an estimate of the average annual consumption in testing. To get this estimate, especially if it be glorified by a phase like "optimum test size," one must know what questions the tests are supposed to answer and how frequently. This in turn means "getting into the skull" of the JCS. We must assume that first of all there is sufficient reason to conduct the tests, even at the risk of compromise of properly-classified information. We know that there will be a finite inventory, and that testing reduces that inventory, whether or not it be formally divided into tactical and non-tactical portions. We can then ask the question: how does the result of an additional test change our perception of the system reliability, and so of the sufficiency of the lesser inventory of missiles to conduct a military mission should it be committed to combat at a future date? Possible answers are discussed in Chapter V. As there are circumstances under which the answer is insensitive to the size of the inventory, we shall spend more time considering the case where inventory for test has no tactical mission.

A long string of heads or tails when flipping pennies is not impossible or even incredible; but after some number, one is entitled to wonder if the coin is biased. Similarly, when testing a missile which is alleged to have high reliability, a string of failures--even a short one--challenges the presumption; contrariwise, a long string of successes tends to be uninformative. In either case there is a practical limit to the value of the additional information in an outcome merely extending such a string.

To address this problem we shall invoke the discipline of Sequential Analysis, to include Sequential Probability Ratio Tests and test series truncation. Much of this is "old hat", having been developed in World War II, most notably by Abraham Wald (Reference 11) working on military problems, and largely standardized by now. It has recently been reported that the methods were independently developed simultaneously by Alan Turing while working at Bletchley Hall to crack the German ENIGMA codes (Reference 12). More importantly there is recent substantive new work not yet "codified" in text books. Two applications of sequential analysis to the Pershing missile test problem will be presented: one by Nozer Singpurwalla and Robert Launer (Chapter III) and one by Michael Woodroofe (Chapter IV). While aspects of the treatment will appear more "frequentist" than Bayesian, both evolve into completely Bayesian solutions. In this paper I shall extract from their work, and comment on it as appropriate. The author of this memorandum is not by profession a statistician, and so requests that the original researchers not be blamed for errors in translating their work into this format.

Chapter III

Launer and Singpurwalla's Proposal

The following submission by Launer and Singpurwalla is the product of over a year of research by the authors, initiated and guided in discussions with the writer of this note. I believe it successfully addresses the problem placed before the authors. Note that all the appendices to this article are to be found at Appendix E.

As the numerical example in the following exposition employs fictitious data and arbitrary values of the parameters α , β , and ∇ , the numerical results should not be taken as applicable to the Pershing II problem. The dependencies and the savings from sequential analysis are however clearly indicated, the penalty when tests are batched, and the potential for squeezing information out of small samples. The next chapter reports further steps toward savings through careful test design.

MONITORING THE RELIABILITY OF PERSHING II MISSILES--A CRITIQUE OF THE CURRENT METHODOLOGY AND A SUGGESTED COMBINED BAYESIAN-SAMPLE THEORETIC APPROACH +

Ъу ∙

Robert Launer* Nozer D. Singpurwalla**

1. INTRODUCTION, TEST REQUIREMENTS, AND ASSUMPTIONS

The reliability of the Pershing II missile arsenal is an unknown parameter which presumably could change over time. To monitor the reliability, and also to ascertain the amount of change in reliability, if any, a sample of n Pershing II missiles is chosen from the arsenal every year, and each missile fired to observe its success or failure. The testing is destructive, and the arsenal inventory is not replenished. Thus, it is highly desirable to <u>reduce</u> the number of test missiles fired year after year. Also, if possible, it is desirable to have the total number of missiles fired per year be a multiple of three-that is, 3, 6, 9, etc. A stated requirement with respect to the year by year detection of change in reliability is that a change of \triangle should be detected with a probability of π or more. Since the test data are

⁺ The authors' appendices are incorporated in this paper as Appendix E. DW

^{*} U.S. Army Research Office, Research Triangle, N.C.

^{**} George Washington University, School of Engineering and Applied Science, Washington, D.C.

of a pass-fail nature, a correct probability model for describing them is the binomial.

Our goal is to determine a sample size and a decision criterion that will satisfy the above requirement, and minimize the total amount of testing. Since each missile is expensive to produce and test, there is a keen desire to incorporate into the analysis <u>all knowledge that is</u> <u>available</u>, both, from the previous tests and engineering experience. Thus a Bayesian point of view is natural here.

2. CRITIQUE OF PRESENT METHODOLOGY

Based on our reading of the pertinent literature that has been made available to us, and our discussions with several people familiar with the test, it is our understanding that the current methodology for analyzing the Pershing II data is based on Fisher's exact test, henceforth FET. We claim that this technique is <u>inappropriate</u> for the situation described above. Furthermore, a modified version of the FET which has been used in similar situations is not appropriate, either. Whereas the FET can be used to detect the equality or otherwise of two binomial populations, it is not designed to detect a <u>specified difference</u> between the two binomial parameters in question. Furthermore, FET <u>does not address</u> the key question of sample size selection, and thus fails to answer the main question posed by our problem. A choice of the sample size should be based on an assumed or target value of the reliability, and this is nowhere apparent in the test.

Given a sample size and the test results from this sample, the FET can give us the "p values" for deciding upon the difference or

otherwise of the two binomial populations in question, and this may be the sole motivation for using this test here.

3. THE COMBINED BAYESIAN-SAMPLE THEORETIC APPROACH PROPOSED HERE

Our proposed approach addresses the issues posed before, and attempts to do this in an economical manner with respect to sample size.

Since reliability changes over time, we introduce an index t, where t = 1, 2, ...; thus t = 1 denotes the first year of testing, t = 2 denotes the second year of testing, and so on. Let n_t denote the number of missiles to be tested in time period t; n_t is the (unknown) sample size, one of our decision variables. Let x_t denote the number of missiles that fire successfully in time period t; note that $0 \le x_t \le n_t$.

Let p_t be the chance that any missile fired at t will fire successfully, or its propensity to do so. Since p_t is unknown to us, we express our uncertainty about it by a probability distribution, say $g(p_t \mid previous failure data, if any, and H)$. Thus p_t is treated as an unknown parameter, and the vertical line in $g(\cdot)$ denotes conditioned upon or given, and H denotes our background information about p_t . If we have no previous failure data, then $g(p_t \mid H)$ denotes our prior <u>distribution</u> for p_t ; otherwise $g(\cdot \mid \cdot)$ denotes our posterior distribution.

If for each time period t we judge the missiles in the arsenal to be <u>exchangeable</u> (we have here finite exchangeability), then it is appropriate to assume that <u>given</u> p_t , the probability of observing x_t

successful firings in a sample of size n is a binomial distribution; that is,

$$P\{x_{t} \text{ successes in } n_{t} \text{ firings } | p_{t}\} = \begin{pmatrix} n_{t} \\ x_{t} \end{pmatrix} \begin{pmatrix} x_{t} \\ p_{t} \end{pmatrix} \begin{pmatrix} n_{t}^{-x} \\ t \end{pmatrix} \begin{pmatrix} n_{t}^{-x} \\ p_{t} \end{pmatrix} (1 - p_{t})^{n_{t}^{-x} t}$$
(1)

The choice of the sample size n_t is based on the following sample theoretic arguments for testing hypotheses about p_t .

If p_t , the chance that a missile is fired successfully at time t, is large, then the number of failures in a sample of size n_t would tend to be small. Given an n_t and having specified a p_t , let x_t^* be the largest integer for which the chance of observing x_t^* or fewer successes is small, say α ; that is,

$$P\{x_{t}^{\star} \text{ or fewer successes in } n_{t} \mid p_{t}\} = \sum_{j=0}^{x_{t}^{\star}} {n_{t} \choose j} p_{t}^{j} (1 - p_{t})^{n_{t}^{-j}} \leq \alpha .$$
(2)

If p_t were to change to $p_t - \Delta$, with Δ large, then the number of failures in a sample of size n_t would tend to be large; if Δ were small, the number of failures in n_t would tend to be small. Thus, for some small number β ,

 $P\{x_t^* \text{ or fewer successes in } n_t \text{ firings } | (p_t - \Delta)\}$

$$= \sum_{j=0}^{x_{t}^{\star}} {n \choose j} (p_{t} - \Delta)^{j} (1 - p_{t} + \Delta)^{n_{t}^{-j}} \ge 1 - \beta.$$
⁽³⁾

If in (2) and (3) we assume that p_t , α , β , and Δ are the only known quantities, then (2) and (3) can be simultaneously solved to obtain an n_t and x_t^* . Once this is done, (2) can be used to test the null hypothesis that the reliability of the missile arsenal at time t

is P_t , with a Type I error α . This is done by accepting (rejecting) the null hypothesis whenever $x_t > (\leq) x_t^*$, where x_t is the total number of successfully fired missiles in a sample of size n_t . If $\alpha = .25$ and $\beta = .25$, then (3) assures us that n_t and x_t^* are suitable for detecting the desired changes in reliability. Note that (3) describes the <u>power of the test</u> as specified by (2), for changing values of Δ . If the null hypothesis is accepted, we conclude that the reliability of the missile arsenal at time t is p_t .

In our case p_t is not specified, as it is an unknown parameter which is liable to change over time. What we have instead is

- i. a prior distribution for p_t at time (t-1), say
 - $g(p_t \mid (n_1, x_1), (n_2, x_2), \dots, (n_{t-1}, x_{t-1}), H), t \ge 2$ and $g(p_1 \mid H);$

ii. a posterior distribution for p_t at time t, say

 $g(p_t | (n_1, x_1), \dots, (n_t, x_t), H), \text{ for } t \ge 1$.

Thus, if we uncondition on p_t , (2) and (3) would become

$$\int_{0}^{1} \int_{j=0}^{x_{t}^{*}} {n \choose j} p_{t}^{j} (1-p_{t})^{n} t^{-j} g(p_{t} \mid (n_{1},x_{1}), \ldots, (n_{t-1},x_{t-1}), H) dp_{t} \leq \alpha,$$

for $t \ge 2$, and

$$\int_{0}^{1} \sum_{j=0}^{x_{t}^{*}} {n \choose j} p_{t}^{j} (1 - p_{t})^{n_{t}^{-j}} g(p_{1} \mid H) dp_{1} \leq \alpha , \text{ for } t = 1 ; (4)$$

$$\int_{\Delta}^{1} \sum_{j=0}^{x_{t}^{*}} {n \choose j} (p_{t}^{-\Delta})^{j} (1 - p_{t}^{+\Delta})^{n_{t}^{-j}} g(p_{t} \mid (n_{1}, x_{1}), \dots, (n_{t-1}, x_{t-1}), H) dp_{t}$$

$$\geq 1 - \beta , \text{ for } t \geq 2 ,$$

$$\int_{\Delta}^{1} \sum_{j=0}^{x^{\star}_{t}} {n \choose j} (p_{t}^{-\Delta})^{j} (1-p_{t}^{+\Delta})^{n_{t}^{-j}} g(p_{1} \mid H) dp_{1} \ge 1-\beta , \text{ for } t=1.$$
(5)

6

In order to obtain the pair (n_t, x_t^*) , for $t \ge 1$, we need to solve (4) and (5) simultaneously. Note that a solution to (4) and (5) would depend on our choice of $g(p_t | \cdot)$. If for example, $g(p_t | \cdot)$ is a member of the family of beta density functions, then (4) and (5) would involve incomplete beta functions and would call for numerical methods for solving them. A method for undertaking this is described in Appendix A. A computer code for implementing the method of Appendix A is given in Appendix B. An example using the above is in Section 5.

As an alternative to the above, and one which is easy to implement, we replace p_t in (1) and (3) by \tilde{p}_t , the modal value of $g(p_t \mid (n_1, x_1), \ldots, (n_{t-1}, x_{t-1}), H)$. The modal value is the most likely value of p_t , given all the previous data, and is determined by the prior distribution $g(p_t \mid (n_1, x_1), \ldots, (n_{t-1}, x_{t-1}), H)$. The posterior distribution $g(p_t \mid (n_1, x_1), \ldots, (n_t, x_t), H)$ represents our best assessment of the arsenal reliability at time t, given all the data up to and including that obtained at t. Its model value \hat{p}_t could be used as a single number which describes p_t . In the next section, we discuss an implementation of the above alternative procedure. An implementation that in computing the pair (n_t, x_t^*) p_t is not replaced by the modal value of its prior distribution.

and

3.1 Assessing Our Uncertainty about p and Procedure Implementation

Since p_t can take values between 0 and 1, a convenient but flexible way for us to express our uncertainty about p_t is via the family of beta density functions on (0,1). Thus,

> 1. We start off our assessment and monitoring procedure by assigning a prior distribution for p_1 , say $g(p_1 | \gamma, \delta, H)$, which for the two unknown parameters $\gamma > 0$ and $\delta > 0$ is a beta density function

$$g(p_1 \mid \gamma, \delta, H) = \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\gamma - 1} (1 - p_1)^{\delta - 1}, \quad 0 < p_1 < 1.$$
(6)

The modal value of the above density is

$$\tilde{p}_1 = \frac{\gamma - 1}{\gamma + \delta - 2}$$
.

Clearly, p_1 best describes in the form of a single number our assessment of \tilde{p}_1 , prior to testing at time t = 1. Furthermore, \tilde{p}_1 is also to be used for determining the pair n_1 and x_1^* , for testing at time t = 1.

- 2. We thus replace p_t by \tilde{p}_1 in (2) and (3), and simultaneously solve these to obtain n_1 and x_1^* . [In Appendix A we discuss how to obtain n_1 and x_1^* without using \tilde{p}_1 , and by directly solving (4) and (5).]
- 3. We take a sample of size n_1 and test these to determine x_1 , the number of missiles that fire successfully. If $x_1 > (<) x_1^*$, we accept (reject) the hypothesis that the reliability of the missile arsenal at time 1 is \tilde{p}_1 .

4. If we accept the above hypothesis, then we update our opinions

about p_1 in light of n_1 and x_1 via the posterior distribution $g(p_1 \mid (n_1, x_1), H)$. The modal value of this posterior distribution is

$$\hat{p}_{1} = \frac{\gamma + x_{1}^{-1}}{\gamma + \delta + n_{1}^{-2}}$$
,

and this number best summarizes our assessment of p_1 after testing at time 1. We now go to step 5.

- 5. If the aforementioned hypothesis is rejected, our choice of γ and δ needs to be revised. This should be done following a more detailed analysis about p_1 . We then go back to stage 1.
- 6. The posterior distribution $g(p_1 \mid (n_1, x_1), H)$ now serves as the prior distribution for p_2 , and its modal value \hat{p}_1 is set equal to \tilde{p}_2 . Thus

$$\tilde{p}_2 = \frac{\gamma + x_1 - 1}{\gamma + \delta + n_1 - 2}$$
,

and p_t is now replaced by \tilde{p}_2 in (2) and (3), which are solved for n_2 and x_2^* . [In Appendix A we discuss how to obtain n_2 and x_2^* by directly solving (4) and (5).]

 We now repeat the steps 3 through 6, and continue the above procedure. Thus, at time (t-1) we have

$$\hat{p}_{t-1} = \tilde{p}_t = \frac{\gamma + x_1 + x_2 + \dots + x_{t-1}}{\gamma + \delta + n_1 + n_2 + \dots + n_{t-1}^2}$$
(7)

as our single best assessment of the reliability of the arsenal at time (t-1), after observing the results of the test at

time (t-1). It also represents our choice for p_t in . equations (2) and (3), for determining the sample size n_t and the decision variable x_t^* .

8. Suppose that at time t, we test n_t items, observe x_t successes, and based on this result, reject the null hypothesis that $p_t = \tilde{p}_t = \hat{p}_{t-1}$. Then we conclude that the reliability of the arsenal has changed from its previous value \hat{p}_{t-1} . When this happens, we investigate the cause for this change, choose some new values, say γ' and δ' , and estimate p_t by

$$\hat{p}_{t} = \frac{\gamma' + x_{t} - 1}{\gamma' + \hat{o}' + n_{t} - 2}$$
.

We now continue as before, bearing in mind that the previous date $(n_1, x_1), \ldots, (n_{t-1}, x_{t-1})$ are no more appropriate for inclusion in our assessment process.

An alternative to the beta prior which has properties of robustness is currently under investigation. However, there is no assurance that the alternative prior will be void of computational difficulties.

3.2 Sequential Sampling to Reduce the Amount of Testing

At any stage t, given an n_t and x_t^* , a further reduction in the amount of missiles tested can be achieved if the testing is done sequentially, one item at a time. Specifically, we would test one item at a time; and stop the test as soon as x_t the number of successes is larger than x_t^* . Thus, ideally, the number of missiles tested could be

as few as $x_t^* + 1$; this implies a saving of $n_t - x_t^* - 1$. The maximum of missiles tested would of course be no greater than n_t . The resulting sample size, that is the number of missiles actually tested at each stage is known as a curtailed sample.

For the above scheme, given p_t we can compute $E(n_t | p_t)$ the expected number of missiles tested using standard arguments--these are shown later. However, since p_t is not known, we average out p_t with respect to its prior distribution to obtain $E(n_t)$, the unconditional expectation of the number of missiles tested at each stage under the sequentially taken curtailed sample. This is shown below.

Given n_t and x_t^* , the probability that $n_t = x$, when a sequential sampling scheme is used is

$$p\{n_{t}=x|p_{t}\} = \begin{cases} \begin{pmatrix} x-1\\ n_{t}-x_{t}^{\star}-1 \end{pmatrix} & (1-p_{t})^{n_{t}-x_{t}^{\star}} & p_{t}^{\star} & (n_{t}-x_{t}^{\star}) \\ & (1-p_{t})^{n_{t}-x_{t}^{\star}} & p_{t}^{\star} & (n_{t}-x_{t}^{\star}) \\ & (1-p_{t})^{n_{t}-x_{t}^{\star}} & p_{t}^{\star-(n_{t}-x_{t}^{\star})} \\ & + \begin{pmatrix} x-1\\ n_{t}-x_{t}^{\star}-1 \end{pmatrix} & (1-p_{t})^{n_{t}-x_{t}^{\star}} & p_{t}^{\star+1} & (n_{t}^{\star}-x_{t}^{\star}) \\ & + \begin{pmatrix} x-1\\ x-x_{t}^{\star}-1 \end{pmatrix} & (1-p_{t})^{x-x_{t}^{\star}-1} & p_{t}^{\star+1} & , & x_{t}^{\star} < x \leq n_{t} \end{cases}$$

In order to obtain $P\{n_t=x\}$, we average out the above by $g(p_t|\cdot)$, where

$$g(p_t|\cdot) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} p_t^{\gamma-1}(1-p_t)^{\delta-1}$$

When the above is done, we have

$$p[n_{t}=x] = \begin{cases} \begin{pmatrix} x-1\\ n_{t}-x_{t}^{\star}-1 \end{pmatrix} \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} & \frac{\Gamma(x-n_{t}+x_{t}^{\star}+\gamma)\Gamma(n_{t}-x_{t}^{\star}+\delta)}{\Gamma(\gamma+\delta+x)} \\ for \quad n_{t}-x_{t}^{\star} \leq x \leq x_{t}^{\star} \\ \begin{pmatrix} x-1\\ n_{t}-x_{t}^{\star}-1 \end{pmatrix} \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} & \frac{\Gamma(x-n_{t}+x_{t}^{\star}+\gamma)(n_{t}-x_{t}^{\star}+\delta)}{\Gamma(\gamma+\delta+x)} \\ + \begin{pmatrix} x-1\\ x-x_{t}^{\star}-1 \end{pmatrix} \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} & \frac{(x_{t}^{\star}+1+\gamma)\Gamma(x-x_{t}^{\star}-1+\delta)}{\Gamma(\gamma+\delta+x)} \\ for \quad x_{t}^{\star} \leq x \leq n_{t} \end{cases}$$

from which $E(n_t)$ can be computed. The above formula can also be used to plot a histogram of the various values of n_t , for each stage t.

If the sequential tests are to be done in batches of 3 rather than testing a single item at a time, the savings in the number of items tested will be less. However, this is still better than compulsarily testing all the n_t items. We do not have a general formula like (9) above to compute the expected sample size. The calculations will have to be done on an enumerative basis. These are shown in Appendix C.

4. COMMENTS ON THE PROPOSED APPROACH

The proposed approach is a combination of sample theory and Bayesian statistics. The former is used to determine the sample size, and the latter is used for inference about p_t . One may express reservations about a procedure in which two philosophical viewpoints are used simultaneously. However, upon closer examination of the approach, such a concern should be dispelled, since the sample theory approach is not used for making inferences about p_t ; it is used for choosing a sample size. The selection of the sample size after averaging out p_t with respect to its distribution $g(p_t | \cdot)$, see equations (4) and (5), makes our analysis fall under the category of what is known as <u>pre-posterior</u> <u>analysis</u>, a perfectly legitimate device within the Bayesian paradigm [cf. Box (1982)].

The monitoring of p_t is done within the Bayesian framework, and besides "coherence" it has the advantage of inducing economy by virtue of the fact that all our relevant previous data are incorporated into the analysis. Furthermore, it allows the incorporation of any engineering or judgmental knowledge that we may have about the missiles into our analysis -- this is done via the parameters γ and δ or γ' and δ' , etc.

5. APPLICATIONS TO DATA

Our proposed approach is designed to specify a sample size for testing at each stage, and thus its effectiveness cannot be fully appreciated if we apply it to existing data. However, we shall apply it

to some given (sanitized) success failure data to demonstrate the fact that the computations of Appendix A can be undertaken, and to compare the results of our main procedure and the simplified alternative, described in Section 3.1. In Table 1, we present the given success failure data, our Bayesian estimate of the mode of p_t at each stage using a uniform prior distribution at stage 0 updated at successive stages using failure data, and the values of x_t^* and N_t using the main procedure and the alternative.

A few facts emerge from an examination of Table 1.

- A large number of items to be tested is called for, when the prior is uniform, with mode .5.
- 2. The number of items to be tested is the smallest when the mode of p_r is closest to 1, namely, at .9.
- 3. The number of items to be tested under the main procedure is always equal to or larger than that under the alternate procedure. This is because the alternate procedure puts all the probability mass at the mode, whereas the main procedure disperses the probability mass over [0,1], with a concentration at the mode.

5.1 Results of Curtailed Sequential Sampling

The sequential sampling approach discussed in Section 3.2 was applied to the data and the results of Table 1. The n_t and the x_t^* values considered were those given by the "alternative procedure"; this procedure gave us smaller values of the n_t 's than the main procedure.

TABLE 1

Results for Main Procedure and Alternative, Using Sanitized Data, and Assuming a Uniform Prior at Stage O

Data		Mode of p	Computed Values of x* and n t t				
			Main Procedure		Alt. Procedure		
Success	Failure		x* t	nt	x* t	ⁿ t	
		.5 00	2	29	5	17	
6	0	.875	8	13	9	13	
11	1	.900 、	10	14	8	11	
11	1	.906	11	15	8	11	
11	1	.909	8	11	8	11	
9	3	.875	9	13	9	13	
9	3	.853	10	15	8	12	
8	4	.825	9	14	9	14	
4	0	.833	11	17	9	14	
3	2	.820	10	16	9	14	
9	0	.837	10	15	9	14	
8	1	.841	10	15	10	15	
7	2	.836	10	15	9	14	
9	0	.848	10	15	8	12	
7	1	.850	10	15	8	12	
	Da Success 6 11 11 11 9 9 9 8 4 3 9 8 4 3 9 8 7 9 8 7 9 7	Data Success Failure 6 0 11 1 11 1 11 1 11 1 11 1 9 3 9 3 8 4 4 0 3 2 9 0 8 1 7 2 9 0 7 1	Data Mode of pt Success Failure .500 6 0 .875 11 1 .900 11 1 .900 11 1 .909 9 3 .875 9 3 .853 8 4 .825 4 0 .833 3 2 .820 9 0 .837 8 1 .841 7 2 .836 9 0 .848 7 1 .850	Data Mode of Pt Main Pr Success Failure .500 2 6 0 .875 8 11 1 .900 , 10 11 1 .900 , 10 11 1 .900 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 10 11 1 .909 , 8 9 3 .853 , 10 8 4 .825 , 9 4 0 .833 , 11 3 2 .820 , 10 9 0 .837 , 10 8 1 .841 , 10	Data Mode of Pt Computed Value Success Failure Mode of Pt Main Procedure \$\success\$ 500 2 29 6 0 .875 8 13 11 1 .900 \ 10 14 11 1 .900 \ 10 14 11 1 .900 \ 10 14 11 1 .900 \ 10 14 11 1 .900 \ 11 15 11 1 .909 8 11 9 3 .875 9 13 9 3 .853 10 15 8 4 .825 9 14 4 0 .833 11 17 3 2 .820 10 16 9 0 .837 10 15 7 2 .836 10 15 9	Computed Values of x_{t}^{*} Mode of p_{t} Main Procedure Alt. Pr Success Failure Mode of p_{t} Main Procedure Alt. Pr Success Failure Success Success	

14

: ~

The expected sample sizes when testing is sequential, in batches . of 3 as well as one item at a time, were computed. These are shown in Table 2. The advantage of testing one item at a time is clear from an inspection of columns 2 and 3 of Table 2.

We also note the <u>overall reduction in sample size</u> using the approach of this paper. The expected sample size can be as small as 9.

The detailed calculations leading us to Columns 2 and 3 of Table 2 are given in Appendix C.

6. PROPOSED FUTURE WORK

An objectionable feature of the proposed procedure, from a Bayesian point of view, is the testing of hypotheses about \tilde{p}_t using the decision variables x_t^* , t = 1, 2, ... The proper Bayesian way to study this problem would be via a <u>Kalman filter</u> model which contains two unknown states of nature, p_t and m_t , where m_t denotes the <u>drift</u> in p_t . The Kalman filter would not only have the ability to monitor the reliability of the arsenal, but would also provide us with a vehicle for <u>predicting</u> the future arsenal reliability. The following are our ideas on how a Kalman filter model for this problem can be developed.

Let Y_t denote some transform of x_t/n_t , and one which makes Y_t approximately normal. The <u>observation equation</u> for the Kalman filter model would be

$$Y_t = p_t + \gamma_{lt}$$

where γ_{lt} is a disturbance term with mean 0 and variance σ_{lt}^2 . We can postulate the following as system equations:

 $p_t = m_t + \gamma_{2t}, \text{ and}$ $m_t = m_{t-1} + \gamma_{3t}.$

TABLE 2

Expected Sample Size for Curtailed Sequential Sampling in Batches of

Stage t	Expected Sample Size for Batch Size 3	Expected Sample Size for Batch Size 1	x* t	nt
0	11.84	10.91	5	17
1 -	12.03	10.66	9	13
2	10.29	9.45	8	11
3	10.37	9.51	8	11
4	10.40	, 9.54	8	11
5	12.28	11.08	9	13
6	11.07	10.16	8	12
7		11.74	9	14
8	12.79	11.69	9	14
9	12.87	11.78	9	14
10	12.78	11.67	9	14
11	13.59	12.72	10	15
12	12.78	11.68	9	14
13	11.14	10.22	8	12
14	11.14	10.21	8	12
		ł	1	

Size 3 and Size 1.

.

In the above equations, we are saying that P_t , the unknown state of nature, consists of a low frequency drift term m_t , which represents a smooth variation in P_t , and γ_{2t} , which is a high frequency component that represents drastic changes in P_t . We assume that γ_{2t} is a normal variate with mean 0 and variance σ_{2t}^2 . The drift term is assumed constant, except for slight disturbances in it; these are described by γ_{3t} , which is also assumed normal with mean 0 and variance σ_{3t}^2 .

The Kalman filter solution would result in uncertainty statements about p_t and m_t , via their distribution functions. These, of course, would be conditioned on $(n_1, x_1), \ldots, (n_t, x_t)$. Large values of m_t would indicate a drift in the arsenal reliability, and so m_t could be used to monitor the change in the arsenal reliability.

It appears that the Kalman filter solution would have several advantages over the proposed approach. The problem of choosing n_t in the context of a Kalman filter is an open question, and this calls for some basic research, assuming that this has not been done before.

A third possible direction for future research is the development of a sequential procedure for testing the missiles. A sequential procedure employing Bayesian considerations may add a further dimension to this problem.

Chapter IV Woodroofe's Proposal

The proposals of Michael Woodroofe are not yet formally documented, but are contained in a series of letters and lecture notes (References 13-17). In this chapter I shall mostly quote from this material with the author's permission, noting that any published versions may differ markedly from those given here. I accept responsibility, however, for the accuracy of the material quoted and the interpretations and extensions of it.

All of the calculations described in this chapter were carried out by Dr. Woodroofe and/or myself. I have programmed most of them for an HP-41C, and listings are given in Appendix D. Instructions and copies on magnetic cards are available. Dr. Woodroofe has used an Apple computer.

Section 1. (Extract from Reference 14).

The Truncated Sequential Probability Ratio Test.

Illustration with a sequential test of the type of savings which are possible and the loss of information which results from the savings. Note that the process starts with the conventional Uniformly Most Powerful test, to be terminated when a specific number Sn of failures has been observed; or when, out of a planned test of size n, the number of observed successes assures that the number of failures cannot reach Sn; or after n tests if not terminated earlier. The choice of n is at this time arbitrary; the value 12 was used in the example to permit comparison to the Pershing test program, past and planned.

÷.,
/We start with a discussion \overline{of} the problem of sequentially testing /such that/ that a failure probability does not exceed a given level. I will illustrate the type of savings which are possible and the loss of information which result from the savings with a specific example.

Let X₁, ... X₁ be i.i.d. random variables which take the values 1 and 0 with probabilities p and q = 1-p, where 0 , is unknown; and consider the problem of testing

Let $S_k = X_1 + ... + X_k$, $1 \le k \le 12$. Then the (UMP) test which rejects H₀ If and only if $S_{12} \ge 4$ has power function

(1)
$$_{\beta_{0}}(p) = 1 - \sum_{k=0}^{3} {\binom{12}{k}} p^{k} q^{n-k}, \quad 0$$

Of course, it may not be necessary to take all 12 observations to determine whether $S_{12} \ge 4$. The test may be curtalled at time

$$t_{o} = \min\{k \ge 1: S_k \ge 4 \text{ or } S_k \le k-9\}$$

Then

(2)
$$E_{p}(t_{0}) = \sum_{k=4}^{12} k {\binom{k-1}{3}} p^{4} q^{k-4}$$

$$\sum_{k=9}^{12} k \binom{k-1}{8} q^9 p^{k-9}, \quad 0$$

* Identically and Independently Distributed.

** Uniformly Most Powerful.

is the expected sample size of the curtailed test.

Selected values of $\beta_0(p)$ and $E_p(t_0)$ are listed in columns 2 and 4 of Table 1 below.

Observe that the type I error probability is .0922 when p = .15 and the type II error probability is .2253 when p = .4.

I tried to construct a truncated version of the SPRT whose power function matched β_0 as closely as possible. Wald's approximations allow one to match the power function at two points. I picked .15 and .40. Wald's approximations then give formulas for the upper and lower stopping boundaries in the (k, S_k) plane. These are listed in columns 2 and 3 of Table 2. There are two problems with these boundaries: Wald's approximations tend to overestimate the error probabilities; and I wanted the test to take at most 12 observations. After some experimentation with formulas (3) and (4) below, I was led to the upper and lower boundaries listed in columns 4 and 5 of Table 2.

Thus, I considered the sequential test which takes

 $t = min\{k \ge 1: S_k \le a_k \text{ or } S_k \ge b_k\}$

observations and rejects H if and only if $S_t \ge b_t$, where a_k and b_k are as in Table 2.

The power function and expected sample size may be easily computed. Let

$$f_{k}(j,p) = P_{n}\{S_{k}=j, t > k\}$$

for k = 0, ..., 11, j = 0, 1, 2, ..., and <math>0 . Then the power function and expected sample size are

(2)
$$\beta(p) = \sum_{k=1}^{\infty} f_{k-1}(b_k-1, p) \cdot p$$

11

and

(4) $E_p(t) = \sum_{k=1}^{12} k\{f_{k-1}(b_k^{-1}, p) p + f_{k-1}(a_k, p)q\}$

for $0 . Thus, one need only compute the values of <math>f_k$; and this is easy in view of the initial conditions, $f_0(0,p) = 1$ and $f_0(j,p) = 0$ for $j \neq 0$, and the recursion

(5)
$$f_k(j,p) = [p f_{k-1}(j-1,p) + q f_{k-1}(j,p)] |\{a_k < j < b_k\}$$

for k = 1, ..., 12, j = 0, 1, 2, ..., and 0 A</sub> denotes the indicator of A.

(

The power function and expected sample size may be computed from (3), (4), and (5). Selected values are listed in columns 3 and 5 of Table 1.

Observe that the power functions β and β differ by at most .0103 for the values computed. This is much better than I had expected when I began the exercise. Observe also that the expected I sample size of the modified SPRT is substantially smaller than that of I the curtailed test when p is small.

After the test has been performed, one may set confidence limits for p by using the relationship between tests and confidence intervals. Order the possible outcomes in a clockwise manner, as in column 1 of Table 3. For each r, 0 < r < 1, one may test the hypothesis

 $K_r: p \ge r$

as follows: the acceptance region A(r) of the test consists of an initial segment of outcomes, in the order of Table 3; one includes precisely enough outcomes to make

 $P_{r}(A(r)) \ge .90$.

Then, after the test has been performed, an upper confidence bound p for p may be obtained from the relation

 $p \leq p^{\star}$ lff $(t, S_t) \in A(p)$.

This is essentially the approach of Siegmund (1978, Biometrika), but substitutes exact calculations for his approximations.

I list some approximate 75% upper confidence bounds for p in Table 3 These were obtained by linear interpolation with formulas like (3).

To the extent that the modified sequential test takes fewer observations than the curtailed test, one may expect less accurate estimation of p.

	Table 1: Power	Functions and Expe	cted Sample Sizes	
- P	β ₀ (p)	β(_P)	E _p (t _o)	E _p (t)
. 05	. 0022	.0022	9.47	6.93
.10	.0256	. 0251	9.92	7.85
15	.0922	. 0899	10.23	8.62
20	. 2054	. 2004	10.40	9.13
.25	. 3512	.3434	10.31	9.35
.30	. 5075	.4975	10.02	9.30
.40	.7747	.7644	9.00	8.57
. 50	.9270	.9204	7.77	7.42

Here: Column 1 is computed from (1), column 2 from (3), column 3 from (2), and column 4 from 4.

	The SPRT		Kodi	fled
k	a, k	b [*] k	ak	^b k
1	-1	2	-1	3
2	-1	3	-1	3
3	-1	3	-1	- 3
4	-1	3	-1	4
5	0	3	-1	Į,
6	0	4	0	4
7	0	4	0	. L
8	1	4	0	4
9	1	4	1	4
10	1	5	1	4
11	1	5	2	4
12	2	5	3	4

Table 2: Upper and Lower Stopping Boundaries in the (k, S_k) Plane

Here columns 2 and 3 are from Wald's approximations; columns 4 and 5 are ad hoc approximations.

Table 3:	Approximate 75%	Upper	Confidence	Bounds

•

Du	tcome	Confidence Bound
t	s _t	
3	3.	
5	4	01
6	4	. 51
7	Ь	.70
5	1.	.61
0	4	- 55
9	4	• 5
10	4	45
11	4	
12	4	.42
12	2	- 39
,,	5	.34
	. Z	.29
9	1	.21
6	0	

••

-.

.

,

Conments by DW:

As indicated in Chapter III, expectations of $\frac{3}{2}$ and E can be computed based on a prior probability distribution. Closed-form solutions exist for "o and Ep(to) for a Beta prior, among others. For (p), and Ep(t), numerical integration is necessary. Other indices derived from the fk(j,p) in manners like that for $\frac{3}{2}$ or E(t) can also be meaningfully be averaged over a prior distribution. As $\Gamma p(t)$ has here a narrow range of variation, its expectation value will not be very sensitive to the choice of the prior distribution. Section 2. (Extract from Reference 15).

To clarify some of the points raised in Section I, Woodroofe provided a more extensive treatment of the development of the limits on observed successes and failures at which the test is terminated. It begins with the method described by Wald (op. cit.) and then continues with a procedure, somewhat judgmental, for modifying those boundaries to reduce the expected size of the test while retaining its power. 1) Testing $H_0: 0 > .15$ is the same as testing 0' = 1 - 0 < .85. If you want to have

$$P_{\theta}$$
{decide $\theta' > .85$ } < α_{0} for $\theta' < .85$

and

$$P_{\theta}$$
{decide $\theta' < .85$ } < α_1 for all $\theta' > \theta'_1 > .85$,

where a_0 and a_1 are small and .85 < $\theta' < 1$, then you cannot simply reverse the 1 roles of zero and 1 in the test described in my earlier letter. A new test must be constructed. See (2) below.

In <u>/Section 1</u> θ was the probability of a system failure.

2) For testing H_0 : $\theta < \theta_0$ at level a_0 with type II error at most a_1 when $\theta > \theta_1$, where $0 < \theta_0 < \theta_1 < 1$ are specified, the SPRT continues sampling as long as

(+)

 $1/A < L_n < B$

where $B = (1-\alpha_1)/\alpha_0$, $A = (1-\alpha_0)/\alpha_1$, and L_n is the likelihood ratio. One finds

$$L_n = \exp \{\Delta_1 S_n - n \Delta_0\}$$

where

 $\Delta_1 = \log \theta_1(1-\theta_0) - \log \theta_0(1-\theta_1)$

$$\Delta_0 = \log (1-\theta_0) - \log (1-\theta_1)$$

and $S_n = X_1 + \ldots + X_n$, n > 1.

Since S_n are integer valued, equation (*) may be rewritten

$$a_n < S_n < b_n$$

$$a_n = \left[\frac{1}{\Delta_1}(n\Delta_0 - \log A)\right]$$

$$C_4 = \left[\frac{1}{\Delta_1}(n\Delta_0 + \log B)\right] + 1$$

where [x] is the greatest integer which is less than or equal to x.

Suppose now that one wants the test to be truncated at M say. Then one wants boundaries a_n and b_n , $1 \le n \le M$. What I did in the example was the following. Let a_M and b_M be such that

 $a_M \leq a_M = b_M - 1$ and $b_M \leq b_M$,

say two integers near the middle of the interval from a_M to b_M . Then let

$$a_n = \max \{a_n, a_M - (M - n)\}$$

 $b_n = \min \{b_n, b_M\}$

and

(::)

for n < M. This gives a first approximation to the boundary. In the example, I then computed the power function of the sequential test with boundaries a_n and b_n and compared it with the power function of the fixed sample size test. I then changed a few of the boundary points to get better agreement between the two power functions. The adjustments were minor and tended to make the continuation region fatter.

The reason that you can't pin me down on the adjustments is that it is a trial and error operation.

$$P_{\theta}\{t=k, S_{k} = b_{k}\} = f_{k-1}(b_{k} - 1; \theta) \cdot \theta$$

and

$$P_{\theta}\{t=k, S_{k} = a_{k}\} = f_{k-1}(a_{k}; \theta) \cdot (1-\theta)$$

Then $P_{\theta}\{\overline{X}_{t} > x\}$ is the sum of these probabilities over all pairs (k,a_{k}) and (k,b_{k}) for which $a_{k}/k > x$ or $b_{k}/k > x$.

4) For inverse sampling there is just one boundary. For curtailed sampling, there are two. Let

 $t^+ = \min\{k > 1; S_k > 4\}$

and

 $t^{-} = \min\{k > 1: k - S_k > 9\}$

Then

$$E_{\theta}(t^{+}) = 4/\theta$$

and

$$FA(+-) = Q/(1-A)$$

The stopping time for the curtailed fixed sample size test is

 $t_0 = \min(t^+, t^-)$

ν So

$$E_{\theta}(t_0) < \min\{E_{\theta}(t^+), E_{\theta}(t^-)\}$$

When $\theta = .15$, $E_{\theta}(t^{-}) = 10.6$.

The formulas for $E_{\theta}(t^{+})$ and $E_{\theta}(t^{-})$ hold for all $\theta_{1} \ 0 < \theta < 1$.

5) I think of the boundaries as a modified S.P.R.T. In the example, they were similar to the curtailed fixed sample size test, but sufficiently different to reduce the expected sample size by about 1 over the range of interest.

6) The calculations in my letter to Launer are for fixed θ . To do . a Bayesian calculation, one would average them over θ values

The formulas which I gave for computing the power and expected sample implicitly assume that that the boundaries a_n and b_n are non-decreasing in n.

Section 3 (Extract from Reference 16).

The Truncated SPRT, Aggregated over Several Tests.

Derivation of a conservative estimate of the probability that in 10 years of testing, at 12 missiles planned for expenditure each year, no more than, say 100, will be needed using the proposed stopping rules. This is to explain how savings in expected sample size may be translated into savings of units which must be purchased prior to the experimentation. For definiteness, I illustrate the method with the truncated SPRT, which is described in / Section I/

In particular, recall the computation of

$$E(k,j;p) = PR(T>k,S_k=j),$$

where p denotes the true failure probability, S_k denotes the number of failures after k units have been tested, and t denotes the stopping time. From this, one gets

$$G(k;p) = Pr(T < k) = 1 - \sum_{j=0}^{k} f(k,j;p)$$

and
$$g(k;p) = Pr(T=k) = G(k;p) - G(k-1;p)$$

for k = 1, ..., 12 and 0 .

Suppose that the truncated SPRT is run n times, say once each year for n years, where n is a positive integer. Then there will be a sequence p_1, \ldots, p_n of unobservable true failure probabilities and a sequence t_1, \ldots, t_n of random sample sizes. Here I regard p_1, \ldots, p_n as unknown parameters, and suppose that t_1, \ldots, t_n are independent random variables for which

$$Pr(t_i=k) = g(k;p_i)$$

for k = 1, ..., 12 and i = 1, ..., n. If $p_1, ..., p_n$ are really random variables, then the calculations described below are valid, if the conditional distribution of $t_1, ..., t_n$ given $p_1, ..., p_n$ is as just described.

Let T denote the total number of units used during the tests,

 $\mathbf{T} = \mathbf{t}_1 + \dots + \mathbf{t}_n$

Then the distribution of T is required. The distribution of T is the convolution of the individual distributions of t_1, \ldots, t_n . This depends on p_1, \ldots, p_n in a complicated manner, but it is possible to find the sharp bound which is valid for all possible choices of p_1, \ldots, p_n . That is, it is possible to find a function H for which

Pr(T k) > H(k)

for all k = 1, ..., 12n and all possible choices of $p_1, ..., p_n$.

I describe the derivation below. The values of H are included in Table 2 in the special case that $\neg n = 10$. Observe that then

Pr(T > 105) < .054

for all p_1, \ldots, p_n . The bound is reasonably sharp, since Pr(T>105) = .050 when all of p_1, \ldots, p_n are equal to .27.

While the bound is sharp, the approach is conservative, since it ignores data from previous years and assumes the worst possible values for p_1, \ldots, p_n . If an independent verification is required for each year, then some of this conservatism may be unavoidable.

The derivation of the bound uses the notion of stochastic dominance. If X and Y are random variables with distribution functions F and G, then Y is said to be stochastically larger than X if and only if $G(z) \leq F(z)$ for all z. If X and X' are independent random variables and Y and Y' are independent random variables and if Y and Y' are individually stochastically larger and X and X', then Y+Y' is stochastically larger than X+X' (as is easily verified); and this result extends from two summands to several. To apply this result, let

 $G(k) = \min G(k;p),$

where the minimum extends over $0 . Then, for any choice of <math>p_1, \ldots, p_n$, the distribution of T is stochastically dominated by the sum of n independent random variables having common distribution function G. Computing G is straightforward. For k < 6, the minimum is attained when p = 0 and G(k) = 0. For k > 6, I computed G(k;p) for a grid of p values and took the minimum over this grid. The values are listed in Table 1. I used a grid width of .01.

TABLE 1. Values of G(k;p)

P	k =	6	7	8	9	10	11
. 24		.2313	.2590	.2966	.5032	• 5556	.7723
. 25		.2222	• 2535	÷ 2955	.4967	5537	.7685
. 26		.2144	.2496	. 2962	.4923	• 5538	.7661
. 27		.2081	. 2474	. 2987	.4900	• 5559	:7651
. 28		.2032	:2468	.3030	:4897	. 5599	.7654
. 29		.1996	.2477	.3088	.4914	•5657	.7669
.30		.1974	.2501	.3163	.4949	.5731	.7696
.31		:1964	.2540	.3252	.5002	.5819	.7735
.32		.1967	.2593	.3355	.5072	.5922	.7783
.33		.1981	.2659	.3472	.5156	.6036	.7840
. 34							

.35

Ĵ

Minimum .1964 .2468 .2955 .4897 .5537 .7651 (... Mean and St dev $\mu = 9.4528 - \sigma = 2.1992$ = 12.1 - (a+b+c+d+e+f)

 $= 12 \cdot 1 - (a+b+c+d+e+f)$ $G^{2} = 12^{2} - ((12+ii))f - ((1+id))e - (10+4)d - (9+if)e - (12+if)e - (12+if)e - (10+4)d - (9+if)e - (12+if)e -$

TABLE 2. Values of H

 $\left(\right)$

1

ſ

H(k) - H(k-1)1 - H(k)k .0460 100 .2026 .0404 .1622 101 .0349 .1273 102 .0295 .0978 103 .0244 .0734 104 .0197 .0537 105 .0155 .0382 106 .0118 .0263 107 .0088 .0175 108 .0063 .0112 109 .0043 .0069 110 .0029 .0040 111 .0018 .0022 112 .0011 .0012 113 .0006 .0006 114 .0003 .0002 115

Comments by DW:

Let
$$g(k) = G(k)-G(k-1)$$
. 4.1

Then
$$d(n,z) \equiv \sum_{k=0}^{n} z^{k}g(n-k)$$
 4.2

is a generating function of the distribution g(k). The generating function for the dominant of m years' test results is then

$$\int (n,m) \equiv \left[d(n,z) \right]^m = \sum_{J=0}^{nm} Z^J d_J, say, \qquad 4.3$$

and the dominant of the probability that a specific number J of tests can be forgone is given by the coefficient dJ of zJ in the expansion of D(r,m).

In our example n=12, and the g(k) for k < 6 are all zeros. Sample data are given in Table 3. So, for m=10,

$$D(12,10) = 4.4$$

$$\left[g(12) + zg(11) + z^{2}g(10) + z^{3}g(9) + z^{4}g(8) + z^{5}g(7) + z^{6}g(6)\right]^{10}$$

$$= \left[g(12)\right]^{10} + 10zg(11)\left[g(12)\right]^{6} + \cdots$$

TA	BL	ιE	3
----	----	----	---

```
g(k)
```

	P = .85 Batch Size		P =	.75
			Batch	Size
k	1	3	1	3
12 11 10 9 8 7 6	.2349 .2114 .0640 .1942 .0487 .0504 .1964	.5103 0 0 .2933 0 0 .1964	.0940 .1258 .2235 .1694 .0361 .1549 .1963	.4433 0 .3604 0 .1963

In Woodroofe's notation

$$d_{J} = H(nm - J) - H(nm - J - I).$$

In particular, in our case,

$$d_{0} = H(120) - H(119) = [-H(119) = [g(12)]$$

is the dominant of the probability that all 120 are required (none can be foregone). It follows that

-10

$$e_{J} = \sum_{i=1}^{J} d_{i} = I - H(nm - I - J)$$

is the dominant of the probability that at most J can be forgone; the generating function for eJ is

$$= (\nu, m) = \sum_{n=1}^{\infty} z_1 \sigma^2 = D(\nu, m) \setminus (1 - \overline{z}).$$

The calculation of the dJ or eJ presents no difficulty except possibly in the control of round-off errors for J large. Sample results are given in Tables 4 and 5 partly repeating material in Table 2, with differences presumably due to differences in accuracy between our computers.

In actual conduct of Follow-on Tests, three failures in a row, or two with an identifiable cause, would be sufficient justification for halting the test until the problem were (identified and) fixed. There would then remain some number of missiles from that year's allocation available for intensive investigation of the fault and for demonstration of remediation. It is not clear that any additional missiles would need to be allocated to those missions, as they could serve the FOT mission at the same time.

It is a trivial matter to revise the expression for D(n,m) to treat the case of batched tests: for example, in groups of 3. Tables 3-5 compare the results for single and triple tests. For the data in the example, whatever the number of missiles considered an adequate inventory for 10 years' testing without batching, about 6-10 more would be required when fired in batches of 3. The analysis in Chapter III gave a similar result.

Up to this point the development has assumed that up to 12 would, in fact, be expended if necessary to provide the foundation for an annual confidence estimate. The question now is: why

TABLE 4

P = .85

Singles

..

۰.

Batches of 3

k	d J= H(k)−H(k−1)	eJ= 1-H(k)	dj	eJ	J
120	5.1E-7	5.1E-7	.0012	.0012	0
119	4.5E-6	5.1E-6			1
118	2.0E-5	2.5E-5			. 2
117	.0001	.0001	.0069	,0081	3
116	.0001	.0002			4
115	.0003	.0006			5
114	.0006	.0012	.0224	,0305	6
113	.0011	.0022			7
112	.0018	.0040			8
111	.0029	.0069	.0511	.0816	9
110	.0043	.0112			10
109	.0063	.0175			11
108	.0088	.0263	.0902	.1718	12
107	.0118	.0382			13
106	.0155	.0536			14
105	.0196	.0733	.1291	.3010	15
104	.0243	.0975			16
103	.0292	.1268			17
102	.0342	.1609	.1545	.4554	18
101	.0392	.2001			19
100	.0439	.2441			20

TABLE 5

$$P = .75$$

Singles

...

Batches of 3

k	dJ= H(k)-H(k-1)	eJ= 1-H(k)	dJ	eJ	J
120	5E-11		.0003	.0003	0
119	7E-10				2
118	0E-9 3E 0		0024	0027	3
11/	3E-8		.0024	.0027	5 4
110	1.45-7				5
115	2.5E-/		0100	0127	6
114	2.0E-0 5.0E-6		.0100	.0127	7
110	J.0E-0	0			8
111	1.4E-J 3.0E-5	0001	0284	.0411	9
110	0001	.0001		•••==	10
100	0002	0003			11
109	0002	.0006	.0604	.1014	12
107	.0005	.0011		• - • - •	13
106	.0010	.0021			14
105	.0016	.0037	.1016	.2031	15
104	.0025	.0062			16
103	.0039	.0101			17
102	.0057	.0159	.1401	.3432	18
101	.0082	.0241			19
100	.0113	.0354			20
9 9	.0152	.0505	.1615	.5047	21
98	.0197	.0702			22
97	.0248	.0950			23
96	.0304	.1255	,1578		24
95	.0364	.1618			25

annually? If an annual series should end without clear resolution, as indeed it must occasionally according to the current plans what then? If there is not a clear cause of alarm, there is no need for alarm.

1

Consider a decision to limit the annual expenditure to 9 missiles, while extending the reporting period to cover 12 missiles (the current standard) if uncertainty had not been earlier resolved. In the worst case (all 12-missile series) reports would occur at 16month intervals, or 8 reports in 11 years. Were the JCS to accept biennial reporting as an (occasional) substitute for annual reporting, this would be a technically simple solution.

Section 4 (Extract from Reference 17)

A Completely Bayesian Stopping Algorithm

[This is my suggestion for doing a complete Bayesian] decision theoretic analysis of the stopping problem. On the basis of the preliminary calculations described below, I estimate that this approach would reduce the number of units needed for testing by at least one per year over the savings which may be attained by using a sequential probability ratio test.

The approach requires the specification of a prior distribution and a loss structure. I suggest a possible form for these quantities below; but other choices would yield to similar analyses.

Let p denote the proportion of non-defective items in the population. Let h_1 denote a density on the unit interval, $0 ; let <math>h_0$ denote the uniform density on the unit interval; and consider prior densities of the form

(1)
$$g(p) = w h_1(p) + (1-w)h_0(p)$$
,

where $0 \le w \le 1$ is a prior parameter. Here h₁ may be thought of as the posterior density which resulted from last year's tests, and w is the probability that p hasn't changed during the past year. If p has changed, which it may with probability 1-w, then it is assumed to be uniformly distributed over the interval $0 \le p \le 1$.

Suppose now that one may observe conditionally independent Bernoulli random variables X_1, \ldots, X_L with common success probability p, given p, and let

 $S_k = X_1 + \ldots + X_k$

denote the number of successes. Then the posterior distribution of p, given X_1, \ldots, X_n is

$$g_{k}(p) = w h_{1}^{k}(p) + (1-w)h_{0}^{k}(p)$$
where $h_{i}^{k}(p) = h_{i}(p;k,S_{k}) \propto p^{S_{k}}(1-p)^{k-S_{k}}h_{i}(p)$
and $\int_{0}^{1} h_{i}^{k}(p)dp=1$

Suppose now that a critical level p_0 is given with the following properties: if $p > p_0$, then the population contains enough good items; if $p < p_0$, then the population no longer contains enough good items and corrective action is desirable; and if p is much less than p_0 , then corrective action is necessary. Suppose further that the purpose of each year's test is to decide whether $p < p_0$ or $p > p_0$; and define one unit of cost to be the cost of testing one item. Then the decision problem may be modelled as follows: the possible decisions are 1 to decided that $p < p_0$ and 2 to decide that $p > p_0$; if one decides that $p < p_0$ when, in fact, $p > p_0$, then one loses C_1 units; and if one decides that $p > p_0$, when, in fact, $p < p_0$, then one loses $C_2(p_0-p)$ units. Here C_1 and C_2 are positive constants. C_1 represents the cost of inspecting the entire system; and the ratio C_2/C_1 is determined by the relative importance of the two kinds of errors.

These three elements, the prior distribution, the sampling distributions, and the loss structure, determine an optimal sampling plan, one which minimizes the sum of sampling costs and expected loss to due an incorrect decision. To describe it, first let m denote the maximum number of tests which could be conducted in any given year (e.g. m = 12). Next, let

$$L_1(k,s) = C_1P(p > p_0 | S_k=s) + k$$

and
$$L_2(k,s) = C_2 E\{\max(0,p_0 - p) | S_k = s\} + k$$

for k = 0, ..., m and possible values of s. Thus L_1 and L_2 denote the conditional expected losses for the two decisions, given $X_1, ..., X_k$, plus the cost of observing $X_1, ..., X_k$. If k = 0, then s = 0 and the expectations are unconditional. If sampling is terminated after k tests, then it is optimal to make decision 1 if and only if $L_1(k, S_k) < L_2(k, S_k)$, in which the expected loss due to terminal decision is

 $L_0(k, S_k) = \min\{L_1(k, S_k), L_2(k, S_k)\}.$

Let $p(k,s) = P(X_{k+1} = 1 | S_k = s)$

for k = 1, ..., m-1 and possible values of s; and define L by

 $L(m,s) = L_0(m,s)$

and $L(k,s) = \min \{L_0(k,s),$

(2)
$$p(k,s)L(k+1,s+1) + (1-p(k,s)L(k+1,s))$$

for k = 0, ..., m-1 and possible values of s. Then the optimal sampling plan is to continue sampling as long as $L(k, S_k) < L_0(k, S_k)$, stopping at time

$$t = min\{k>0: L_0(k, S_k) = L(k, S_k)\}.$$

Here L(k,s) is the minimum expected loss plus sampling cost among all sampling plans which take at least k observations.

If h is a beta density, then it is possible to compute L_1 and L_2 as sums of products of p_0 and $(1-p_0)$ times ratios of factorials. I can supply the details, if you are interested. Using these explicit expressions, it is straightforward to compute L by the backward induction (2); and, once L and L_0 have been computed, it is simple to classify the possible outcomes (k,s) as stopping points, points for which $L_0(k,s) = L(k,s)$, or continuation points. Moreover, the stopping points divide thenselves into lower stopping points for which $L_0(k,s) = L_1(k,s)$ and upper stopping points for which $L_0(k,s) = L_2(k,s)$. If the largest (smallest) lower (upper) stopping point is called a_k (resp. b_k), then

 $t = \min\{k > 1: S_k < a_k \text{ or } S_k > b_k\}$

and it is optimal to decide that $p < p_0$ if and only if $S_t < a_t$.

The several tables which accompany this letter describe the optimal sampling plan in a special case in which m = 12, h_1 is a beta density with parameters a = 6 and b = 2, w = 3/4, $p_0 = 3/4$, $C_1 = 60$, and $C_2 = 180$. Here the ratio $C_2/C_1 = 3$ equates the seriousness of deciding that $p < p_0$ when p > p with that of deciding that $p > p_0$ when $p_0 - p = 1/3$; and the magnitudes of C_1 and C_2 were chosen to make it optimal to take up to about 12 observations. I believe that this is consistent with the power and sample size requirements discussed earlier. In a certain sense, these values of C_1 and C_2 are implicit in those requirements.

Table 1 lists the boundaries a_k and b_k of the optimal test. These boundaries are remarkably insensitive to a+b. I got nearly the same values when a = 9 and b = 3. Table 2 lists an ad hoc modification of the optimal boundaries which takes account of the economies of testing items in groups of three. Table 3 gives the posterior probability that $p > p_0$ for each possible outcome, using the adhoc boundaries. It clearly exhibits the following qualitative feature of the test: if the results of the first six tests this year are consistent with last year's results, then further testing is not optimal. Table 4 gives the frequentist properties of the adhoc test, the power function and expected sample size as a function of p. Observe that the maximum expected sample size is substantially smaller than that of the adhoc test; and recall the crucial role of the maximum in determining the number of items which must be purchased for testing.

TABLE	1:	AN	OPTIMAL	BOUNDARY
				the second s

.

<u>Design</u> <u>Parameters</u> :	m=k, $a=1$, b	=2, w=3/4,	p=3/4, C1	=60, C2=180
k	a	k	bl	k
1 2 3 4 5 6 7 8 9 10 11 12	- 0 1 2 2 3 4 5 5 6 7		- 2 3 4 4 5 6 7 7 8 8	
	TABLE #2:	A MODIFIED	BOUNDARY	
k	a	k	Ы	ĸ
1 2 3 4 5 6 7 8 9 10 11 12	- 0 0 0 1 2 3 4 5 6 7		- 3 4 5 5 6 6 7 7 8 8	
TABLE #3:	POSSIBLE OU	ICOMES WITH	MODIFIED	BOUNDARY
k	S	k	P(p≯po	b)
3 6 6 7 8 9 11 12 12 12 10 8 6 3	0 1 2 3 4 5 6 7 8 7 6 5 3		.0251 .0084 .0507 .0813 .1211 .1634 .1185 .1546 .3111 .4543 .5183 .6517 .7450	

	TABLE #4:	FREQUENTIST PROPERTIES	
P	BETA	MEAN	VAR
.05	.9999	3.4575	1.281
.1	.999	3.8288	2.4702
.15	.9983	4.4161	3.5485
. 2	.9903	4.8134	4.5345
.25	.9788	5.4154	5.43
.3	.9582	5.8102	6,2305
.35	.9244	6.3797	6.8348
.40	.8728	6.7887	7.559
.45	. 8000	7.1384	8.1442

.

Comments by DW:

1.1

With this note Woodroofe completes the transition from Wald's classic treatment to a Bayesian approach. The use of a prior probability which is a mix of two hypotheses is in part an attempt to address the criticism that priors can become too sharply peaked, neglecting the potential staleness of old data. One might still ask whether there should be an upper limit to the value of k used in the prior.

The loss functions included in this section are representative, rather than my recommendation. The variable called po in the functions Ll and L2 could have different values in the two cases.

Chapter V

Other Stopping Criteria

A possible argument for small test sizes may arise after all missiles have been bought: any test reduces the potential tactical inventory. The decision criterion is unfortunately not unique. This chapter discusses a few examples.

Section 1. Utility as a Criterion

Let $\phi(p, s, f) \Delta p$ be the posterior probability distribution of p, given s "equivalent" successes and f "equivalent" failures on which to base a prediction. Let U (N,p) be the "utility" of an inventory of N missiles of reliability p. The estimate of the utility of the inventory is then

$$U(N) = \int U(N,p) \, d(p;s,f) \, dp$$

Now perform a test: N goes to N-1; with probability p, s goes to s+1; and with probability 1-p, f goes to f+1.

After the test the utility is

$$U(N-1) = \int U(N-1, p) \left[P\varphi(p; s+1, f) + (r-p)\varphi(p; s, f+1) \right] dp.$$

The criterion is: Is U(N-1)>U(N)?

Examples of utility functions are:

Np (expected targets killed);

-Np(1-p) (uncertainty is reduced);

N-T/P (excess inventory, where T is size of critical target list);

 $T[1-(1-p)^{N/T}]$ (expected damage);

 $T\left[\left(-\left(1-\alpha \beta\right)\left(1-\beta\right)^{b}\right]$ (b=largest integer in N/T; a=N/T-b is the fractional part; this reduces to Np for small N, goes to expected damage for large N).

Clearly there is a similarity between this method and that in Secion 4 of the previous chapter.

Section 2. Information as a Criterion

Another criterion would be the information the decision maker gains from the test about the posterior distribution of p. This would be applicable when no single utility function can be agreed on. An example is the Kullback-Leibler information measure on two probability density functions F1 and F2 (Reference 18):

• -

$$I(F_{1},F_{2}) = \int F_{1}(p) \log \frac{F_{1}(p)}{F_{2}(p)} dp.$$

It can be applied to the current problem by defining Fl and F2 respectively as the posterior and prior density functions for p.

Shannon's information measure S(F1,F2) is the expectation value of I(F1,F2) over the observed values of success and failures.

To illustrate, we may identify F2 with expression 1.6 from Chapter I:

$$F_{2}(p) = p^{s_{1}-1}(1-p)^{f_{1}-1}/B(s_{1},f_{1})$$

and Fl with expression 1.8:

$$F_{1}(p) = p^{s_{1}+s_{2}-1}(1-p)^{f_{1}+f_{2}-1}/B(s_{1}+s_{2},f_{1}+f_{2})$$

so that log F1/F2 is

$$log \frac{F_{1}(p)}{F_{2}(p)} = log \left[\frac{\Gamma(n_{1}+n_{2})\Gamma(s_{1})\Gamma(s_{1})}{\Gamma(n_{1})\Gamma(s_{1}+s_{2})\Gamma(s_{1}+s_{1})} p^{s_{2}}(1-p)^{f_{2}} \right]$$

$$= C + S_{2} \log p + f_{2} \log (1-p)$$

where C is the logarithm of the gamma-function combination in curly braces, all independent of p. Noting that

and letting $\psi(z) = \frac{1}{\Gamma(z)} \frac{d\Gamma(z)}{dz}$, the logarithmic derivative of the gamma function, the expression for I(F1,F2) reduces to

$$I(F_{1},F_{2}) = C - s_{2} \left\{ \Psi(n_{1}+W_{1}) - \Psi(s_{1}+s_{2}) \right\} - f_{2} \left\{ \Psi(n_{1}+n_{2}) - \Psi(f_{1}+f_{2}) \right\}.$$

Consider now the case where s2=n2=1 (a single successful trial). Then a b C D/ 177 . .

$$T_{s} = \log \frac{n_{i}}{s_{i}} - \{\Psi(1+n_{i}) - \Psi(1+s_{i})\}.$$

In the alternative case wher S2=0,n2=1 (a single unsuccessful trial)

$$\frac{T_{F}}{F} = \left\{ \sigma_{S} \frac{n_{i}}{f_{i}} - \left\{ \Psi\left(1+n_{i}\right) - \Psi\left(1+f_{i}\right) \right\} \right\}$$
Shannon information is
$$S = \frac{S_{i} T_{S} + f_{i} T_{F}}{n_{i}} \approx \frac{1}{2n_{i}} + \dots$$

and the Shannon informat

As this never goes to zero (for finite nl), the cost of this information must be balanced against the use made of it.

I have not yet found a way to apply this criterion to the Pershing testing problem.

Chapter VI

Conclusion

I return now to the tasking from the Under Secretary of the Army, as given in the opening of this memorandum. The mathematical methods of sequential analysis proposed here for estimating reliability changes possess a rigor not found in the Army's current method, and make clear the risks in following their prescription. They provide a basis for reducing the size of an annual test and so reducing too the cost of a testing program. Indeed, they even challenge the need for an annual report, and suggest that the interval between reports can be enlarged (e.g., to two years) with no increase in risk to management. They do not, however, encompass a variety of other issues which are fundamentally operational in nature: firings to support training, alternate uses of inventory, system life. These must be the subject of further investigation.

Readers of this report may be disappointed that such very different approaches to the stopping problem have been presented in the foregoing chapters. I observe that such a seemingly simple problem has apparently not been hitherto subject to the scrutiny it deserves, and that it is comforting that two separate investigations have reached similar conclusions.

I see ultimately more promise in the methods proposed in Chapter IV, but would recommend that those of Chapters III and IV be applied ---to Pershing using the best available data so that a refined test program can be determined. In Chapter III is proposed the application, as yet unexplored, of Kalman filtering techniques to this problem. This research merits monitoring, if not support.

Appendix A

References

1. J. R. Ambrose (Under Secretary of the Army), "Operational Missile Testing". Memorandum to the Director of the Army Staff, 30 August 1982.

2. D. Willard, "Reliability Tracking by Sequential Sampling for Binomial Distributions", paper presented to 29th Conference on the Design of Experiments in Army Research and Development, October 1983.

3. Data Collection, Evaluation, and Reporting of Ballistic Missile Test Programs, JCS Memorandum SM-477-75 August 1975 (S).

4. Revised Guidelines for Use in Evaluating Strategic Ballistic Missile Operational Test Programs. IDA Study S-364/WSEG Report 92C, March 1975 (S).

5. JCS Guidelines for Missile Operational Testing and Reporting Memorandum SM-395-84, June 84 (S).

6. Revised Guidelines for Use in Evaluating Missile Operational Test Programs. Revision 92D to IDA Study S-364, February 1984 (S).

7. M. G. Kendall and A. Stuart, "The Advanced Theory of Statistics," Hafner, New York, 3rd Edition, 1969.

8. H. F. Martz and R. A. Waller "Bayesian Reliability Analysis", Wiley, New York, 1982.

9. H. Raiffa and R. Schlaifer, "Applied Statistical Decision Theory", Harvard University, 1961.

10. E. T. Jaynes, "Prior Probabilities", IEEE Transactions on Systems Science and Cybernetics, Vol. SS C-4, No. 3, Sept 1968.

11. A. Wald, "Sequential Analysis", Wiley, 1947. (Dover edition, 1973).

12. Andrew Hodges, "Alan Turing: The Enigma", Simon and Schuster, 1983.

13. M. W. Woodroofe, Unpublished Lecture Notes, October 1983.

14. M. W. Woodroofe, letter to R. Launer, August 2, 1983.

15. M. W. Woodroofe, letter to D. Willard, Novem: 10, 1983.

16. M. W. Woodroofe, letter to D. Willard, December 14, 1983.

17. M. W. Woodroofe, Letter to D. Willard, January 17, 1984.

18. P. K. Goel and M. H. DeGroot "Information about Hyperparameters in Hierarchical Models." Journal of the American Statistical Association, Vol 76, p. 140, 1981.

Appendix B

Bibliography

Mann, Schafer, and Singpurwalla, "Methods for Statistical Analysis of Reliability and Life Data", Wiley, 1977.

C. P. Tsokos and I. N. Shimi (editors), "The Theory and Applications of Reliability", Academic Press, 1977.

J. J. Deeley, M. S. Tierney, and W. J. Zimmer, "On the Usefulness of the Maximum Entropy Principle in the Bayesian Estimation of Reliability, IEEE Transactions on Reliability, Vol. R-19, No. 3, August 1970. Appendix C

Unclassified Extract from Reference 4:

.....

•.

Revised Guidelines for Use in Evaluating Stategic Ballistic Missile Operational Test Programs.

IDA Study S-364/WSEG Report 92 C, March 1975(S)

W366- 926, Cint,

D. ARABASIS MEDILES ANY

(U) The variable characteristic queue in the transfer of the enjorace to endow so the data should be specified. The instruction of the other data processing haveleted as drawing numerical performance estimates from the transfer three threed be of englished for each performance between ediments from the transfer threed in the calculations should be tangential to perform the vertication of the analytical approach.

E. SENGLENITY ANALYTIS

(1) A constituity unalysis should be conducted for each performance estimate to indicate whether the numerical results would change significantly if the transmit of test or data appendies view changed.

F. CONTIDENCE STATEMENTS

(U) Two types of confidence statements should be provided for each performance factor:

- (1) A statistical confidence bound based upon the quantity of data used in computing the factor.
- (2) A qualitative assessment based upon the quality of data used in computing the factor.

The qualitative assessment should be based upon an appraced of the validity and applicability of the test date as outlined in Part 1 of these guidelines.

(U) The statistical significance of differences in estimates of performance factors that is indicated by comparisons of the results of different sets of Operational Test data should be addressed and statistical confidence statements regarding these differences should be provided. The results of one method for comparing reliability samples is illustrated in Table 4.



U. State 1123

۰.

•

7

.

. . . .

:

4

Dariel Charts	Le 5 S.t "B"		Difference in	
Est the type Macrosse Pauloj	No. of Tests	Ruin bi ay (Success Rutin)	Date Sets "A" and "E"	Significanse of Difference in Felixbility1
3 0/50 = 1.00	5	2/5 = .40	03	+63.
	10	4/10 = .40	00	+69.
	15	6/15 = .40	03	+€0.
	5	3/5 = .00	40	.98
	10	6/10 = .00	40	.6≎≁
	15	9/15 = .00	40	.99+
	5	4/5 = .8.0	20	.£3
	10	8/10 = .80	20	.94
	15	12/15 = .80	20	.97
27/30 = .90	5	2/5 = .40	50	.97
	10	4/10 = .40	50	.99+
	15	6/15 = .40	50	.92+
	5	3/5 = .60	30	.86
	10	6/10 = .60	30	.95
	15	9/15 = .60	30	.98
	5	4/5 ≈ .€)	10	.54
	10	8/10 ≈ .80	10	.63
	15	12/15 ≈ .80	10	.69
	5	5/5 = 1.00	+.10	.38
	10	10/10 = 1.00	+.10	.59
	15	15/15 = 1.00	+.10	.71
24/30 = .80	5	1/5 = .20	60	.98
	10	2/10 = .20	60	.03+
	15	3/15 = .20	60	.95≁
	5	2/5 = .40	40	.91
	10	4/10 = .40	40	.98
	15	6/15 = .40	40	.99
	5	3/5 = .60	··.20	.68
	10	6/10 = .60	··.20	.80
	15	9/15 = .60	20	.86
	5	5/5 ≈ 1.00	+.20	.73
	10	10/10 ≈ 1.00	+.20	.85
	15	15/15 ≈ 1.00	+.20	.93

[13] J. C.O. State tile? Statificance of the Difference of Pellubility Data on Two Sets of Test Data

50

UNCLACCIAND

ł;

.
Det. Set "A"*	[] [] []	- Set "p"	Difference in	
Robet aliye (Shurete Robie)	110. of Tests	Fich White (Success Tratio)	R hability Printion Lista Sots "A" and "D"	Toutof Significance of Datesunce in Faliabilityt
21/20 = .70	5	1/5 = .20	50	.95
	10	2/10 = .20	50	.90
	15	3/15 = .20	50	.89+
	5	2/3 = .40	30	.79
	10	4/10 = .40	30	.91
	15	6/15 = .40	30	.99
	5	3/5 = £0	10	.49
	10	6/10 = .00	10	.59
	15	9/15 = .60	10	.74
	5	4/5 = .80	+.10	.45
	10	8/10 = .80	+.10	.57
	15	12/15 = .80	+.10	.63
	5	5/5 = 1.00	+.30	.80
	10	10/10 = 1.00	+.30	.95
	15	15/15 = 1.00	+.30	.99

Table 4 (17) (Compared)

*The number of tests in Data Set "A" is 30 for all cases shown,

TThe values shown (F) are obtained by using Fisher's Exact Text:

$$P = 1 - \sum_{\nu=S_{1}}^{\nu'm_{2}x} {\binom{N_{1}}{\nu}} {\binom{N_{2}}{S_{1}+S_{2}-\nu}} / {\binom{N_{1}+N_{2}}{S_{1}+S_{2}}}$$

where $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x!}{y! (x-y)!}$

$$\frac{S_1}{N_1} > \frac{S_2}{N_2}$$

$$\frac{N_1}{max} = \frac{N_1}{S_1 + S_2}$$
 whichever is smaller

 $N_1 = number of models in sumple set 1$

 N_2 = number of tests in semiple set 2.

 $S_1 = number of successes in sample set 1.$

 $S_7 =$ number of successes in sample set 2

Sec A. Hald, Statistical Theory With Engineering Applications, John Wiley and Sons, Inc., 1960, p. 709.

GERESCHUSTED

Appendix D

HP-41 Programs

The HP-41 handheld calculator is slow but remarkably powerful. For example, a program listing for the standard Fast Fourier Transform (FFT) algorithm is no lengthier than that for a FORTRAN version and because of some quirks of the HP-41, the program is in some ways more efficient. With a 56-bit word, numerical accuracy is higher than in most personal computers, and so round-off problems are slower to arise.

Reported in this appendix are a set of programs written for this study. Their original purposes were to give or to verify solutions, but they have two additional values justifying their inclusion here: they demonstrate that the mathematics called upon is not intractible and can be packaged small, and they may be useful as is to others working the same or related problems.

The first group provde solutions to Equations 1.9 and 1.11 and thus can be considered a proper means of getting the answers wrongly sought via Fishel's Exact Test. The versions given are lengthy but are relatively robust to the accumulation of round-off errors. Included is the program PII, written to be a model for and to verify calculations of Singpurwalla and Launer.

The second group provide handy means of exploring Woodroofe's treatment of sequential analysis. ET provide solutions to Equations 1 and 2 of Chapter III, Sec 1. BND provides Wald's and Woodroofe's boundaries of the region of test continuation; and MW permits computation of a number of properties of a test plan defined by BND. LOP computes boundaries using the Bayesian method of Chapter III, Sec. 4.

Not included is a package of routines which manipulate truncated Taylor series and was used to compute the expansion of D(n,m) given in Eq 4.4. This is available from the author.

The memory requirements of an HP-41CV or CX are needed, and if it is not the CX version, then an Extended Functions module (XF) with its Expanded Memory. The occasional use of Synthetic Programming can be circumscribed, or if the programs are identical to those listed here, they should run on any version of the HP-41 with adequate memory and the XF module.

JCS+ Implements Eq.1.9 and DA+ Eq.1.11.

They call for inputs and report the value of the integral as "CL=" for Confidence Level. The plus sign means there are no subtractions in the algorithm, hence less round-off error.

PII Implements Eqs. 4-6 of Section III.3.

Entering at LBL A leads to an evaluation of \checkmark and at LBL \mathcal{B} to evaluation of β . Lines 51-62 clear a block of registers, using program BC in a module called PPC ROM. This can be replaced by ordinary coding. If Flag 02 is set, then the summation sign in Eq.4 or 5 is ignored; only a single term is considered. Subroutines 1, 2, and 13 are the core of algorithm.

EΤ

• •

Solves Eqs. 1 and 2 of Section IV.1.

$$\beta_{o}(p) = \sum_{k=e}^{N} {\binom{N}{4}} p^{k} {\binom{i-p}{k-k}} and$$

$$E(t_{o}) = \sum_{k=e}^{N} k {\binom{k-i}{c-i}} p^{c} {\binom{i-p}{k-c}} + \sum_{k=N-e+i}^{N} k {\binom{k-i}{N-c}} p^{t-N+c-i} {\binom{i-p}{k-c+i}} = c p^{c} \sum_{i}^{N-c} {\binom{k+c}{c-i}} {\binom{i-p}{k}} + {\binom{N-c+i}{(1-p)}} {\binom{i-p}{k-c+i}} + {\binom{N-c+i}{N-c+i}} {\binom{i-p}{k-c+i}} P^{t}.$$

Calls for N, c, and p (unadjusted values will be used as is).

Memory utilization keyed to that in MW: N, c, and p in same registers.

MW

Requires two files in Extended Memory named Am and Bm where m is a number provided in response to query "FILE#?" or is already stored in register 19. (Routine BND may have been used to create these files.)

Start program at line 1 or at LBL E; line one to provide/revise the value of N, the maximum number of tests. At E, provide "p" and "FILE#." If RAD-DEG selection set to RAD, program computes and reports G(k) as required by Section IV.3; if set to DEG, this is ignored.

Program reports β (p), E(t), and a (p) (which in effect interchanges meaning of "reliable" and "unreliable". Sect IV.1.

LBL B produces output stating "bi/i = cumulative probability of sufficient failures to halt." Accumulates probability of exit passing clockwise around boundary. If there are several points on boundary at N=N max, then these are labeled F. Then program continues along "a" boundary.

LBL C does the same as LBL B but counterclockwise.

LOP

To meet the goals of Section IV.4. Computes the boundary conditions for continued testing, based on the loss functions L1 and L2 (which can have associated with them different criteria P1 and P2, as well as cost factors C1 and C2).

Program invites all necessary input insertion/revision/ verification, and then constructs a diagram of the operating space. To conserve space this pattern is stored as packed binary data (a la flags). LBL J provides a visualization of this pattern, for display or printing (see figures below). This algorithm has also been run on a Commodore for verification.

Routines 6 and 7 support generation of loss functions L; and L2 If others are chosen, these must be rewritten along with some of Routine 2 (lines 57-100).

BND

Develops the boundaries to be used in MW, by Wald's and Woodroofe's methods. Input called for: PO, Pl, a, and b (later, m). 0 < PO < PI < 1. Level of test = a. Probability of Type II error = b (P≥ P1). Ho: $p \leq po$. (Section IV.2). M is number of tests. Wald's methods, and and bn' reported out. Lines 1-85: 86-156: Woodroofe's modification. 157-END: Subroutine E. Calls for a file number k; then stores Woodroofe's boundary numbers an and bn in files AK and BK. If Flag 25 is clear to start, program halts if attempt is made to overwrite existing file. Set the Flag to permit overwriting.

A1+LBL "UCS"	5141 RI 01	96+LBL 03
02 CF 29	52 RCL 01	97 RCL 11
03 "DEL="	57 STO 87	98 RCL 06
64 SF 00	33 310 61	99 YtX
85 .	5441 PI - 92	100 ST* 12
A6 XER AB	55 DP1 02	101 RCL 02
07 "N1="	53 RUE 05 52 DOI 07	102 E
AR F	JO RUL 07	193 -
80 L 89 XEO 80	J/ - Fo LOCTY	104 RCL 08
07 MEA 00	JO LHOIA	195 +
19ALRI P	37 E	186 LOSTY
10*202 0	6M -	107 YFG 94
11 51-	61 /	100 CT+ 10
16 6 17 950 86	62 RCL 10	100 01+ 12 109 Dri Ai
13 XER 00	63 RCL 07	107 KUL 01 110 E
44-101-0	64 -	110 C
14+LBL U	65 LASTX	111 -
15 "N2="	66 RCL 89	112 KUL 03
16 3	67 +	113 +
17 XEQ 00	68 /	114 LHSTX
	69 ×	115 XEV 04
18+LBL P	70 RCL 80	116 ST/ 12
19 "\$2="	71 /	117 °CL=*
20 4	72 E	118 FIX 4
21 XE9 00	73 X(> 13	119 ARCL 12
	74 *	120 AVIEW
22+LBL 18	75 ST+ 13	121 STOP
23 •REL DEG"	76 196 07	122 RTN
24 AVIEW	77 GTO 82	
25 RCL 00	78 RCL 88	123+LBL 00
26 CHS	79 689	124 FIX 0
27 E	80 PCI 86	125 FS?C 88
28 +	81 -	126 FIX 4
29 STŪ 11	82 LOSTY	127 ARCL IND X
30 RCL 04	02 LHUIN 07 E	128 PROMPT
31 F	00 E 04 -	129 FSPC 22
72 +	04 -	138 STO IND 7
77 PC: 97		171 PTN
74 -	00 KUL 00 07 DC: 11	•••
75 STO 85	87 KUL 11	132+1 RE 94
76 CTO 86	88 /	177 049
30 JIO 00 77 LOSTY	89 \$	174 9759
DILNOIN 70 C	90 KUL 13	175 CICk
30 E 70	91 X(> 12	133 31GN 132 V/N 1
37 - 40 CTO 00	92 *	130 A\7 L 177 CT. V
40 510 66	93 ST+ 12	137 314 1
41 E	94 ISG 86	
42 -	95 GTO 01	138•LEL 83
43 RUL 02		139 X=Y?
44 +		140 GTO 86
45 STG 89		141 ST* L
46 LASTA		142 D3E X
47 CHS		143 GTO 85
48 RCL 01		
49 +		144+LBL 85
50 570 10		145 PEN
		146 B 1
		147 FT4
		148 .550.

JCS+

.

.

 \cap	_1_
H	T

01+LBL *DA+-82 CF 29 03 SF 00 94 "DEL=" **0**5 . 96 XE9 88 07+LBL A 08 "H1=" 89 E 10 XEQ 00 11+LBL B 12 "S1=" 13 2 14 XE9 00 15+L8L 0 16 "N2=" 17 3 18 XEQ 80 19+LBL D 20 *\$2=* 21.4 22 XE9 88 23+181 10 24 "ABS DEG" 25 AVIEW 26 RCL 86 27 1/8 28 E 29 -30 STO 89 31 ROL 01 32 RCL 02 33 -34 STO 11 35 RCL 03 36 + 37 E 38 -39 STG 05 48 RCL 84 41 RCL 03 42 --43 E 44 + 45 STO 86 48 . 47 370 15

48+181 B1	101
49 RCL 86	102
50 STO 07	103
51 RCL 03	195
52 X()Y	186
53 t E4 CTO 10	197
34 510 10 55 100TV	168
56 E	109
57 +	110
58 RCL 05	111
59 +	113
60 STO 12	114
Stat Ri 82	115
62 RCL 12	116
63 RCL 07	117
64 -	118
65 STO 13	120
66 RCL 02	121
67 E	122
69 CHS	123
78 STO 83	124
	120
71+LBL 03	126
72 E	128
73 KUL 02 74 -	129
75 RCL 08	130
76 -	131
77 LASTX	132
78 E	133
79 -	135
80 / 01 BCI 18	136
82 RCL 08	137
83 -	(
84 LASTX	138
85 RCL 13	137
86 X<>Y	141
87 -	142
89 *	143
98 RCL 89	144
91 *	140
92 E	147
93 X() 14	148
74 ₹ 95 ST+ 14	149
96 ISG 62	156
97 GTO 03	
98 RCL 06	
99 RCL 07	
160 -	

91	LAST	X	
32	Ε		
93	-		
84 Se	1		
80. 97	RUL .	11 07	
55 27	RUL	0/	
97 90	1001	·y	
90 R9	RCi	12	
10	XO	, -	
11	-		
12	1		
13	*		
14	RCL	0 9	
15	*		
16	RCL	14	
17	X⇔	15	
18	*		
19	51+	15 47	
20	136	0(0)	
61 22	910 201	92 85	
22 27	ENS.	00	
24	REL	06	
25	-		
26	LAS	TX	
27	Ε		
28	-		
29	1		
30	RCL	09	
31	1		
32	RCL	15	
33	-X⊖	15	
<u>04</u> २८	ж ст.	12	
30	190	10 06	
30	100	00 01	
1	410	~.	
38	+LBL	84	
39	RCL	88	
40	RCL	02	
41	Ε		
42	-		
43	YtX		
44	-5T*	15	
45	RUL	99	
46	CHS		
147 120	Ľ		
140 21 û	т • р г:-	p٩	
172 50	ruu Yti	60	
	110		

151 ST# 16 152 RCL 01 153 E 154 -155 ROL X 156 RCL 82 157 E 158 -159 -160 XEQ 85 161 ST* 16 162 RCL 05 163 RCL X 164 RCL 03 165 E 166 -167 -168 XEQ 05 169 ST/ 16 170 FIX 4 171 "CL=" 172 ARCL 16 173 AVIEW 174 BEEP 175 STOP 176 RTN 177+LBL 05 178 CHS 179 XOY 180 SIGN 181 XO L 182 ST+ Y 183+LBL 06 184 X=Y? 185 GTO 07 186 ST* L 187 DSE X 188 GTO 06 189+LBL 07 190 RDN 191 X(> L 192 RTN 193+L51 00 194 FIX 8 195 FS?C 00 196 FIX 4 197 ARCL IND X 198 PROMPT 199 FS20 22 200 STO IND Y 201 RTH 202 EKP

· •				j <u>i</u> t≜_t_ st	
	TIT			12: XEE 21	
L.	F II	f	2 · · · · · · · · · · · · · · · · · · ·	101 00 GI	
		97 117HIT	<u> </u>	122 Mil 17	
10	$\Gamma \cdot 0$	· · · · · · · ·	Blog Clear Harth Hart	123 RCL 16	
	tinax			104 V4V	
C.	C C	OCTO NOY		127 160	
	Driver		64 FET 66	125 RGL 10	
		og rtn	25 2.1	102 501 67	
	•		0.2 4 7 1	125 KUL 60	
	1		· · · · · · · · · · · · · · · · · · ·	127 RCL 14	
	, .	27 LBL 09	27 VEC 00	100 1	
		08 R) (3)	DI ALE VZ	120 +	
10			68 A85	129 YtX	
		095401	29 STO 89	178 4	
		10 23 03	07 010 07	100 -	
-				131 5!* 13	
		11.5×12	7ê•i R! - ê7	172 PCL 94	
		12 家小 63		102 KOL 04	
—			21 RUL 19	133 E	
		13 617 16	72 INT	174 -	
		14 ROL 17	15 400 TT 070 47		
		(E	73 510 17	135 KUL 16	
		10	74 RCL 89	136 XEQ 84	
		16 E °	95 V_35	100 004 17	
		17	(J X=0:	137 214 13	
I -		··· ··· ··· ···	76 GTÚ 16	138 RCL 65	
-		18 STU 17	77 ENTER4	170 E	
-		19 PTN		137 E	
⊨ €		14 - 141-1 66 - 1971)	78 CHS	140 -	
1 ``		ZØ KIN	70 E	1 11 DOL 12	
-			(7 L	141 KUL 10	
		ALLER A	88 +	142 XEQ 04	
		ZITER H	Q1 STO 10	127 \$7/ 17	
		22 SF 86	01 010 10	140 01/ 10	
		07 670 16	82 /		
		23 610 10	88 OT2 78	144+1 R! 17	
_			00 010 00 04 E		
		OXALDE P	84 - E	145 XEW 19	
.		ZHVEDE D	85 RCL 15	146 RCL 12	
		25+LBL PII	67 I	1/0 KOC CC 447 CT/ 17	
1 X 37		26 CE AA	86 T	147 517 13	
-		20 6: 00	87 STO 14	148 FIX 4	
1.1			00 10077	140 500 80	
		27 ♦ ! R! 16	DO LEDIA	197 FU/ 00	
			89 RCL 16	150 -tb=1	
		28 CF 01	00 DC: 07	151 500 80	
		29 F1% 2	YO NOL OO	131 63: 00	
	. 🖊	52 OF 00	91 +	152 "1-a="	
		30 LF 22	Q2 STO 84	157 F	
C 1 ¹		31 "DEL="			
-		75 10	93 +	154 RCL 13	
		32 18	94 STO 85	155 FC2 81	.
-		33 XEQ 00	0E		
			95.	156 -	
			96 STO 13	157 ARCE X	
		34+LBL U	97 DCI 16	(EQ at Vea	
		35 *CAMMA=*	27 KUL 10	106 F A-	
		74 7	98 STO 88	159 FIX 0	
		35 3	99 ¥±0 ?	150 001 17	
		37 XEQ 00	22 D-V. 400 DTC 00	100 KCL 11	
		70 +75170	166 610 56	161 HEUL A	
		JO DELIN-		162, 28	
		39 15	total D ¹ - G1		
-		AR YED BR	IDIVLDL DI	163 +	
• •		10 524 00	102 XEQ 21	164 X<>Y	
			187 PCI 16	ICS CTO INT Y	
		41+LBL D		100 510 100 1	
-		40 E14 6	184 E	166 HVIER	
5		42 FIA 0	185 +	167 ISG 19	
		43 "N="	407 DOL 64	101 100 17	
~		44 16	106 KUL 14	168 GIU 07	
		10 AF 10	107 RCL 04	169 FS2 81	
-		45 XEW 00	100 001 00	170 VED 00	
L			IDO KUL OV	110 ALW 07	
		12 - 1 - DI - F	109 ST- T	171 BEEP	
1		46€LBL L	110 QT+ 7	175 CTAP	
		47 FIX 0	110 U.F.C.	112 210	
		40	111 ST- Y		
-		70 A-	112 🔹	1774! R) I	
		49 17	4 4 T	AND AN AN	
		50 YEO 80	113 /	174 26.62	
		JO ALK DO	114 +	175 ROL 19	
1_		51 ROL 17	(15 DC) 60	170 NON 17	
	_	52 £07 14	113 KUL 07	176 186	
	(OL RUL IV	116 /	177 +	
	1	53 E	117 CT+ 17		
1		54 -	110 DIA 10	13 ock VILLAN ING ARUT CENT	
			118 DSE 00	179 RTN	
		77 ES	119 GTG R1		
		56 /	117 010 01		
		57 ECN 40		•	
-		과신 비한 · 투입 문제 · · · · · ·			
		58 X V			
I .					

• ,	181 8164	241+LBL 83	284 XEE 84	766 • ,8, 28	1
*	182 X - L	242 ROL 07	305 STZ 13	367 REL 16	1
	183 X=00	243 PCL 01	306 GTO 17	368 E	ļ
	184 670 86	244 ST- Y		369 +	
•-	165 3403	245 /	307+LBL 19	370 RCL 01	
		246 4 347 571 57	308 PCL 03	371 ST- Y	
	186+LBL 85	249 FUL 165 597	309 E	372 /	l l
•	187 ST# L	250 /	318 STO 11	373 Rt	•
	188 BSE X	249 E.	311 -	374 🔹	
	189 **	200 t Ari nor ai	312 STO 82	375 🔹	
	190 DSE Y	201 105 01 050 670 07	313 X=0?	376 E	
	191 GTO 85	202 610 60 057 674 10	314 G10 92	377 +	
		203 314 14	315 KCL 89	378 DSE 01	
	192•LBL 86	2544) RI 14	316 X=87	379 GTO 88	
	193 RUN	2547EDE 14 255 DCL 12	317 640 92	380 RCL (
	194 X(> L	250 KCE 12 256 ST4 17	318 KUL 13 710 F	381 CHS	
	195 KIN	257 RTN	317 E 720 A	382 E	
	1964) RI - 21		520 4	383 + 704 Dr: 12	
	197 601 16	258+LBL 16	321+1 BL 91	309 KUL 10 705 V4V	
	198 PCL 60	259 CF 01	322 ENTERT	200 HA 704 4	
	199 -	268 RCL 16	323 ENTERT	306 + 707 ¢765	
	200 STO 07	261 E	324 RCL 08	200 010: 200 010:	
	201 RCL 17	262 STO 13	325 *	789 F	
	282 \$272	263 +	326 RCL 02	798 +	
	203 X<>Y	264 STO 04	327 /	391 END	
	204 STC 01	265 RCL 03	328 E		
	205 E	266 E	329 X(> 11		
	286 ST+ 87	267 -	330 *		
	207 RCL 04	268 STO 85	331 ST+ 11		
	208 RCL 00	269 RCL 16	332 RDN		
	209 -	270 RCL 15	333 ISG X		
	210 STO 06	271 +	334 **		4
	211 RCL 83	272 510 86	335 DSE 02		۰.
	212 E	273 KUL 17	336 GTO 91		
	213 STO 12	274 310 66			
	214 -	2(3 A-0) 972 pto 19	337+LBL 92		
	215 STO 82	276 610 10	338 KUL 09		
	216 X=0?	97741 DI 17	339 CHS		
	217 GTU 15	2779 DC1 04	340 E 741 -		
	643-4 9 -33	279 P/L 87	391 † 740 DCL 14		
	218+LBL 02	280 RCL 06	342 RUL 14 747 DCL 07		
	219 KUL 00 200 KCL 07	281 RCL 00	343 KUL 63 744 m		
	220 KUL 00 001 DCi 84	282 ST- T	044 4 785 VAV		
	221 KUL 00	283 ST+ Z	746 DCI 11		
	222 RUL 02 227 RT- T	284 ST- Y	747 ±		
	220 01 1 224 ST- 7	285 *	348 STO 12		
	225 ST- Y	286 /	349 RTN		
	226 +	287 *			
	227 /	288 ST* 13	350+LBL *PI*		
	228 *	289 E	351 STO 81		
	229 RCL 08	290 ST+ 13	352 RCL 15		
	230 *	291 DSE 00	353 RCL 03		
	231 ST+ 12	292 GTO 13	354 ST+ Y		
	232 E		355 XOY		
	233 ST+ 12	293+LBL 18	35 6 🗸		
	234 DSE 02	294 RUL 14	357 RCL 89		
	235 GTO 02	295 KCL 16	358 -		
		296 + 667 - 6675	359 STO [
	236+LBL 15	277 LHUIA 900 VCD 07	368 ENTERT		
	237 RCL 01	270 AEW 84 000 87# 17	361 CHS		
	235 X=07	177 017 18 700 DDI 05	362 E	•	
	239 610 14	JOC KUL OJ 70: Dê: Qê	353 + 724		
	240 E	001 KUL 00 1807 4	304 / -/5 5		
		1 OUL 1	(2월주) 1		

Output of MW

A:.::: -2T-	
02 "N="	
8 3.	
04 XEV 00	
#3 *c=*	
00 E	
81 YFA 82	
08 TP=T 60	
ΰÿ Ż .	
ÎN XEV NA	
	•
IZ KUL 00	
13 KUL 01	
14 510 84	
i5 -	
16 STU #3	
17 5	
18 KUL 82	1
19 -	
20 XEW 01	ł
ZI RUL DZ	
27 VIV	
23 116	
24 -	
25 KUL BI	
20 * -	
27 510 85	
20 RCE 00	
27 KUL 01	
30 E	
<u> 31</u> -	
32 510 83	
33 -	
Tá STÚ MÁ	
75 576 57	
3J KUL 02	
JO YEA DI	
37 E	
38 RUL 02	
<u>3</u> 7 -	
AN REL NO	
AI PEL DI	
AT NOL OI	
92 -	•.
43 E	
44 +	
45 TTX	
46 LHSTX	•
47 ¥	
10 T	
77 514 83	
20 .F1=.	•
3 1 APCL 03	
52 HYIEN	

ET

۰.

53	üf i	iō
54 55	Ē	ŪŪ
56 57	+ 570	Má
58 58	LHST	X
59 68	ŘŪL -	81
61 62	STU JE	0 3
63	ŘŨĹ	ÚŻ
64 65	- LAST	X
66 67	/ XF0	Ri
68	RÜL	02
70 70	KÜL TTX	00
7i 72	* 376	ñá
73	DUF	·)=-
74 75	4 XE¥	6 9
76	ŘĨN	
77	LBL.	60
79 Hi	ut 2 III I	C NUX
80 81	PRUI ESZI	ir î 1 22
82	iŭ i	ND T
83	KIN	
841 85	ENTE ENTE	ŪÌ Kt
86 57	ENTE	RT 24
88 88	Ē	
871	, LBL	ÜŻ
90 91	¥ Říli	ñá
72	KÜL	U 3
73 74	85? +	89
95 96	FC?	0 9
97 97	LHST	X
70 77	/ ¥	
109 181	Ē	
102	DSE	0 3
192	LIÚ.	M2 -

LSLB	$\begin{array}{l} n_{3} / i = 0, 0 0 0 0 \\ 3 / 2 = 0, 0 0 0 0 \\ 3 / 3 = 0, 0 0 3 4 \\ 4 / 4 = 0, 0 0 3 4 \\ 4 / 5 = 0, 0 0 3 4 \\ 4 / 5 = 0, 0 0 3 4 \\ 4 / 5 = 0, 0 0 3 5 4 \\ 4 / 5 = 0, 0 0 3 5 4 \\ 4 / 7 = 0, 0 3 5 2 \\ 4 / 1 = 0, 0 3 5 2 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 4 / 1 = 0, 0 5 0 5 \\ 5 = 0, 5 2 2 5 \\ 5 = 0, 5 2 2 5 \\ 5 = 1, 0 0 0 0 \\ 5 = 1, 0 0 0 0 \\ 2 = 1, 0 0 0 0 \\ 1 = 1, 0 0 0 0 \\ \end{array}$
LBL C	a-i/i=0.0000 -i/2=0.0000 -i/2=0.0000 -i/3=0.0000 -i/5=0.0000 0/6=0.377i 0/7=0.377i 0/7=0.377i i/9=0.6224 i/10=0.6224 i/10=0.6224 i/10=0.9303 i=0.9303 i=0.949i i0=0.940i i0=0.940i

BND

Constructs.	BI+LBL -BND-	30 RÚL 07	104+18L 02	To stare istelling F
and	02 -P0-	5i E3	105 REL IND 13	and by 138 -FL #1
disda_c	0 3.	52 /	105 8(1?	Calls (r 139 PRUNPT
c las	04 FIX 4	53 E	107 SF 00	m - 100 STU 11
Doundaires	05 XEV 00	54 +	108 417	Am BM IDI KUL 14
for more	06 -11-	33 JU 08 (ι.∞A)	109 8(2)	IDS. Free 102 -H-
THW, MX.	UT L	56 510 11	110 XEV 04	163 XEV 03
There 0-1.000	NA YEA NN	57 KLL 84	III DSE X	164 RUL 13
- 1961	07 Tar 15 T	38 LH5	ii2	165 -B-
ancarac	10 2	27 210 67 Lung	113 DSE 08	
computed using	11 YEA 00	60 -a -	114 GTU 02	166+LBL 83
Wallteren hier	12 0	61 H510 10	115 -b-	167 RUL 07
a construction of	IS S IA TEU MA	57 570 12 Rad	116 ASTO 10	168 ARCL II
a the war	17 ALA 40		117 RUL 07	KAPATS INY LAFLU
Woohrenter mal	S. 16 FTI H2	AT RUL OU	118 510 88	if despterte 170 LF 23
(-, 10 KCL VC	SS ENCEPT	119 ST+ X	Unist 5=25171 .
(manalicities 33)	18 RU 83	67 ENTERT	120 ENTERT	to ever - ITC SEERTIN
Subrontene E	19/ A	58 XF9 M	121 ENIEKT	er 450 173 X() Z
stores a. 26.	20 LN	69 RLi ii	122 19	E, esnya 114 1.001
•••===• • • • •	21 510 84	70 STÚ 03	123 +	unuint #, 173 #
weld derives	22 E		124 510 13	176 X.21
a. 261	23 RUL 03	72 F	120 +	100 -
	24 -	73 +	126 510 12	178 E
I S N S M.	25 RUL 02	74 STU 19 (lub)+1	127 510 15	1/9 +
Werdricke Hum	26 / B	75 -0 -	128 DH=?"	180 SHYERS
picki and by 3	27 LN	76 ASTO 10	127 FRUNE1 174 STA 188 12	131 KIN
4 16 - + -1	28 'STU 06	77 20	130 510 IMU 12	
- <u>-</u>	27 E	78 RCL 07	131 UN 80	132*LDL 81
Der = pric	30 RUL 00	79 +	1724) B. 47	163 KUL 01
a.46'20.46	<u> 31</u> -	80 STU 12R6	ISZVLUL DO ITT PIL IND IT	104 T 135 CTA 13
	3 2 E	81 KUL 03	100 KOL 160 10	Raphane First Till
an=mulan-	MAR 33 ROL 81	82 ENTERT	135 SE MA	INT TZ LIST E
	54 -	83 ENTERT	136 8217	Mas-[X] isa mui
Pu=muspu, p	M) 35 /	A SA ENTERT	137 8021	atestler / 189 -
(n< M).	36 LN	WALD 85 XEY BI	138 XEN 04	
_	37 5TU 05 Ac	N. 4-1 80 -3-	139 DSE 03	191 LL N
0 < P0 < P, <.	1 38 RUL 01	14 V 87 HSTU 18	140 GTU 03	192 ARCI 10
	39 RUL 89	88 RUL 07	IAL RIN	193 ARUL 00
a = level of HS	40/	89 510 03		174
5=pr. 1 tyxl	arright LN	90 ENTERT	142+LBL 84	195 AFUL X
(P » P,)	$\mathbf{q}_{2} + \mathbf{\Delta}_{1}$	91 ENTERT	143 STU IND 12	196 RUM
	43 511 8 (19,17)4)	92 ENIERT	144 ÜLĤ	197 AVIEN
Ho pr Pu.	44 51/ 85 A-A	93 19	143 ARCE 10	198 156 12
	43 511 11 (Lizixa,	99 t	145 HRUL 05	199
	45 F1X 8	30 510 13 MUN	147 THET	200 156 03
	47 10	76 1	145 AFÚL X	201 ບໍ່ໄປ ບໍ່ເ
	90 (45 -25 45	9(+ 51 175 17 Ø	147 HTIE=	202 RIN
	97 AEV 09	98 510 12 Ram	150 FCPC 00	
		77 319 19	iji kom	203+161 00
		100 カボービー・	152 DBE 13	204 CF 22
		191 ERVER1 191 ITA 195 -	(5)	201 11=
		102 DIV 100 12 127 72 22	124 115 15	200 APOL INC.
		193 UE 85	155 **	201 11-1
			175 R.S.	283 F5 10 L
				105 3 A. C. A.
				19 EN

•••

	シシ				
しいとうという		• 3 • 3		131 101 43	112.21
	91°LEL 77	16 191•C+			
	96 LF 23	40 KUL 14U 1	29 KSL 01 21 E		178 51+ 11
Law (Day and Sal	83 -M=-	4C A71 C	71 L		iry stu ind m.
in the to Beden in		40 AC/1	- 36		וּאָשָׁ אַרָר וּאַ
	03 ATH 00	NUN YE	73 AV15		+ - 7.
(1) ((1) To voi to	86 RUL 88	JU 136 F	74 A(21		
うち たん きょうちょう	ήζĒ.	51 610 BI	73 510 CI	139 E3	
	- 19	52 570 BI	36 E3	149 /	141 CO
· []]]]]]	NY 510 13	13 E	37 /	+ 1.4	189 E .
			78 510 i4		* CD
	1		33 ULA		1 40
			I'MN AC VE	143°LBL 14	187 51+ 12
				I AN I IN I	เรีย ไว้น ผ้า
		2/ L9165		+ 1.5	jag uru nz
	13 - 12- 01	ב. בייבייי אכ	1014-202 00		I 30 KUL IT
	2 E	39 570 IND L	ומכ אור כו		
	ij XER BB	be Ril 15	193 RUL 19		
		bi E3	INI PUI	CHJ REI	
			1 63 -	149 E	
			inh 510 13	1 Ju +	I'YA AUL WS
				i <u>j</u> i -ú-	193 57# 1MI 1
					NUX QUI
	ZO FILETS	A1 D16 C9			141 KI HY
	ži iž	66 510 11		133 HKLL 20	
	ZZ FIK Ö	65 510 12	+ 411		
	23 KEU UD		111 370 87	133 FIA #	177 KUN
		BRFLBL BZ	ILZ RUL IND OF	136 HKUL A	CNU EJ
		61 20 61 20	113 RUL 13	137 HYLEN	/ 192
			ile Ateri		4 202
				i Jue Lui de	283 5TU 14
	21 KLL 00				
	+ 97	C ENIERI	110 KLL 110 00		ZHAFLBL MJ
	+ 67	73 ENIEKT			PH THI THE THE
	58 510 83	74 RUL 04			
	31 LHSTA	+ E	117 KLL 07	1010 201	
	<u>3</u> č +	76 510 87	izo rul oj	IOJ KUL INU Å	
	<u>55</u> 510 86	77 E	121 F3? 83	104 KLL 02	21 112 047
	34 F512E	+ <u>4</u> <u>1</u>	izz ülk	165 *	61 DC1 607
	-Ĥ- 55	79 510 17	123 51= IND T	166 510 16	CA 119 A12
	Jo KUL DA	BO KUN	ich use T	167 51+ 18	•
	37 RUL 03	Fi RUL 83	IZ3 RUL IND T	IOS STU IND BE	211. THE DO
	JE KEE LO	8Č +	izb říl ðz	iéj kül öb	212 21 212
	<u>5</u> 7 - 8-	83 310 BB	izt fl? ob	ITO RUL IND IT	CIS BLL
	ië KCL ëJ	ي. هو.	izë tlă	i7i +	
	41 KUL BO	+ 29	- i <i>č</i> i -	172 E	-=(1)0_ (17
	AZ AFA 20	Rb 510 18	130 5T+ 1ND 03	173 +	
	AJ RUL IND R	10× 12	131 150 19	174 510 17	
		XB X=4	132 GTU 83	ITS RUL INU A	
	•				

346 150 2 1			348 136 1	349 610 09	530 KIN		· 3310161 11		·	534 B3E 1	333 610 87	336 .ENU.																																	
		383 610 80	306 LHSTA	307 8471					312 E	513 +	GIA RT	315 -F-	316 XEW 09	517 X()T		318°LEL B 8	319 XCM	320 ŘÚL Đồ	32i ž	<u>3</u> 22 -	זַבַזַ אַנוּ שָּש	324 ŘČL Z	323 -D-	JZ6 JF 00		327•LBL 83	326 FIX 0	JZY KUL IND Z	338 +	331 F37 BI	332 GTU IB	333 GETA	334 HRUL X	333 RUN	336 -F/-		337+LBL IÐ	338 HRUL T	339 - F=-		341 HRUL K	TAT III			343 619 11
	201 CU3	205 KUL 18	263 +	ZOH YY=OZ	Z63 610 87	KICH I HHY		5 00 C	+ 697	ZTO KT	271 -F ⁻	ZTZ XER BY	Z73 X()T		ZTA+LBL BT	273 XVIT	בֿדָּה אַנוּ פֿז	277 2	Z78 -	ŽĨJ ŘŮL ĐĐ	ZB0 KUL Z	281 -a-	Zaz keu by	ZUJ KIN	- 	Žŭ4+LBL Č	-H- 282	ZÃO FIX 0	287 HRUL IY	ZĂĞ ŬĹĂ	289 SEEKPTH	ZYB KUF	291 RCL 84	ZJZ RŪL ĐĐ	233 E3	294 /	233 E	źÿb +	. 297 .	298 -a-	233 KEU 03		201 KCL 10	382 KLL 11	- 540
•••		CCD ALW DD	221 - F(T)=-		6 1 1			Y THI THIR CZZ	ŽŽĠ PŘUNET	227 F370 22	ZZA STU IND T	223 RIN		ZJÖ•LBL ZÖ	231 HRUL 13	č3č .	233 SEEKPTH	ZJ4 SIGN	233 57+ 2	236 -	237 EJ	238 /	239 +	ZAN GETRA	ZAL KIN		ZĄZOLBL B		6 11 44 Z	Z43 HRUL IY	ZAD ULA	Z47 SEENFTH	ZAB XUF	Ž 45 Řůl 03	בֿשָּׁ אַנוּ שָּשָ	2 <u>5</u> 1 E3	, žčž	233 E	234 +	. <u>667</u>	- 10- - 10-2	201 AEM 07		233 35 81	ŽEÐ ŘĽL 17
· conterdations																																													

٠

•

Lit C i Lit B num

the chersen services on have any og an > > >

LOP

			160 FCL 01
54. D - 61.5E+	54+LEL 01	- 106 LASTA	161 E
81*LDL LDF	55 \$CL #1	109 /	162 +
82 "6"	50 MCL 01 52 CTA 00	119 -	
63 CLX	35 510 62	110	163 /
04 XEQ 90		III KUL GZ	164 +
85 141	Compute 57+LEL 82	112 E	165 STO 15 8 (4,5)
00 Q 02 7	L, L. 58 RCL 83	113 -	166 -1
	¥ 59 STO 87	114 *	167 21 00
W7 XEU UO	will be por at	115 ECI 81	101 452 02
03 °b*	4 00 KUL 04	110 662 01	168 +
B 9 4	61 510 86	110 +	169 RCL X
10 XEQ 80	62 XEQ 06	117 E	170 LASTX
t1 •P•	63 RCL 06	118 -	171 *
10 5	64 *	119 RCL 09	172 -2
12 J	65 \$10 13	128 +	177 /
13 XEM 60	66 816 10 66 VEN 87	(21 Rt	110 /
14 "说"	50 ACH 01		174 51+ 1
15-6	67 RUL 00	Store 122 STU IND :	175 X<>Y
16 XEQ 00	6S *	mi(L, 2) 123 DSE 02	176 RCL 02
17 -11-	69 STŨ 14	(notice 124 GTO 82	177 RCL 00
10 17	78 E	- Stuck 125 DSE 81	178 *
10 11	71 ST0 87	fillune > 126 GTO 81	170 574 7
19 XEU 80	70 610 80	127 PC1 80	117 017 6
29 *02*	72 310 63	100 CTO 91	180 LH51X
21 18	73 XEW BE	128 510 61	181 -
	74 E	129 SF 21	182 +
22 ♦ [R! - 88	75 RCL 06	130 RCLFLAG	183 RCL 01
07 CE 00	76 -	Cler (131.	184 RCL 89
	77 *	£1. \132 STO d	195 +
24 FF 75 75 197 9	78 PCL 13	133 X()Y	100 014 7
25 HELL IND F	70 4	Restore (174 21 47	100 517 2
26 AVIEK		Ening (134 21.40	187 +
27 FS?C 22	OU RUL II	21-43 (15) STOPLAG	188 510 10
28 STO IND Y	81 ×	Sime (136 RULFLHb	189 RCL IND Y 🐤
29 RTN	82 RCL 01	21-43, 137 510 16	190 RCL IND Y
	83 +	c-2- Clier	191 E
- 1 / 70ALE E	84 E	138+LBL 03	192 ST- 18
Englishand and and	85 -	139 PCL 01	167 861 15
31 20	96 STO 17	140 STA 02	175 KUL 13
32 510 89	00 010 15 07 VEG 67	710 010 02	174 514 1
33 FSIZE	57 ALV C7	1414151 64	195 -
34 RCL 80	86 E	141*LDL 04	196 +
35 ENTERT	89 RCL 06	142 RCL 82	197 +
36 ISG X	90 -	143 RCL 03	198 RCL IND 10
77 ••	91 *	144 E	199 XCY
20 ENTER4	92 RCL 14	145 -	299 9772
30 ENTERN 30 100 9	93 +	146 +	200 ACT
39 156 A	94 BOL 19	147 10STY	
40	05 +	140 001 04	202 3/17
41 *	73 4	140 RUL 01	203 SF IND 02
42-2	96 RUL 81	149 +	204 DSE 02
43 /	97 +	150 RCL 04	205 GTO 04
44 84	98 E	151 +	206 RCL 11
45 4	1 99 -	152 /	207 BEL 81
10 T 42 CTO 11	1 100 FCL 13	153 ROL 06	200 KGE 01 060 A
46 510 11	101 9:372	154 *	
47 +		155 5	ZMY RULFLHS
_ Set-20 48 PSIZE	102 ACCE		210 STÚ INU Y
	Min(L. L.) 183 510 14	106 LHS(A	211 RCL 16
50 E	104 FIL 88	157 -	212 STOFLAG
51 51-11	185 Fil 01	158 FIL 83	213 ISE #1
54 S 44	186-2	153 +	214 611
24 - 1 5 - 27 - 24	ė		
Ker wa	• ·		
In and Discus			こ15 こうして

Reports	217+LBL J		268 RCL 62	319 RCL 02
	218 PCL 88		269 RCL 07	328 ROL 67
Carthurson	219 PC- 11		278 +	321 +
Citis for	222 702 77		271 2	722 5
Viewing &	220 · 001 · 007V		272 -	707 614 1
P. 4.	ZZI LHOIA		277 CTA F	323 314 1
ruling	222 8(3)		273 310 L 374 CF 60	329 KUR
,	223 E3		274 LF 00	325 ST- I
	224 /		273 2-07	326 STO 1
	225 +		276 SF 80	327 CF 00
	226 E		277 E	328 X<>Y
	227 +		278 FS?C 00	329 X<=Y?
	228 STO 81		279 GTŪ 11	338 SF 00
			280 ENTER1	331 Rt
	22941 RI 14		281 ENTERT	332 ES20 88
	2234000 14 270 DCL 83		282 ENTERT	777 CTA 17
	200 KUL 00			774 ENTERA
	231 E3		20741 PL 18	JJ4 ENIERI 775 ENTERA
	232 /		SOUNDEL IN	33D ENIERT
	233 E		204 KUL N 205 +	336 ENTERT
	234 +		285 *	
	235 STO 02		286 RCL 1	337+LBL 12
	236 PCL INP 81		287 E	338 RCL N
	237 STOFLAG		288 +	339 *
	238 CL8		289 RCL [340 RCL [
	239 RCL 6		290 ST- Y	341 /
	DAG FOD INT RT		291 /	747 +
	240 101 100 CC		292 *	747 PCL 1
127,42	241 FT 545 FC5 195 GC		293 +	744 DOL 1
·	292 PU? INU 01		294 INF 1	074 RUL L 745
127,33	245 411		294 DJC (295 CTO 10	390 - 744 -
	244 155 82		273 610 10	346 *
Can put	245 STO 6		001.00E 44	347 FUL T
in 1,5L	ア246 SF 12		296+LSL 11	348 -
new with	247 AVIEW		297 E	349 +
5 2-12	248 186 01		298 RCL 65	350 DSE (
Better:	249 GTO 14		299 -	351 GTO 12
Rall	-> 250 CF 12		300 RCL J	
STO FLACE	251 CLX		301 Y1X	352+LBL 13
DN STE DO	252 RIN		302 *	353 ROL 85
			303 RTN	354 201 1
Con pute best	057411- 00		•	755 F
or the	203VL01 00 954 D01 01	D +	364+LBL 07	752 -
U V	204 RUL 01 OFF DOL 07	Compute	785 F	
	255 KUL 87	Pur juz	706 PCL 65	337 11A 750 Loot:
	256 +		300 FOL 00 767 CT_ V	308 LHSIA
	257 PCL 85		367 31- 1	359 /
	258 +		306 / 305 / TC N	<u>354</u> #
	1 259 2		389 512 5	361 ENP
	268 -		310 FLL 01	5036 to
	261 STÙ 1		311 FCL 67	(-, m) (-)
	262 E		312 +	
	263 FCL 85		313 FCL 03	
	264 ST- Y		314 +	
	24 C 25 C 26		315 E	
	and an		716 -	
	100 110 - 110			
	200 200			

.

•

.

۰.

	Sample at put of Submittie J	
·	$\begin{array}{c} M=12, 60.06\\ a=6, 80.019\\ b=2, 60.019\\ b=2, 60.019\\ b=2, 7500\\ M=0, 7500\\ M=0, 7500\\ C1=66, 0000\\ C1=66, 0000\\ C1=66, 0000\\ C1=66, 0000\\ c1=64, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,$	
	M=12,0809 a=6.66001 b=2,0009 F=0.7560 F=0.7560 H=6.7560 F=0.7560 H=6.7560 Run ** 1 ** 1 ** 1 ** 1 ** 1 ** ** ** 1 * 1 * 1	
⊭=12, 8809 ⇒=5, 8008 5=2, 8008 P=6, 5508 U=68, 8880 C2=189, 8008	**************************************	

(0)00 ۰D :*:*:*:*: C2=1890.0A08 C1=180.0008 t * *:*: P1=0.7580 P2=0.5688 a=6.0000 b=2.0000 1=12, 8<u>0</u>89 H=8.7509 .*:*:*: 1.4.4.4 64 *:*: 5 . .

Other versions of ontopet, making clearer the upper boundary. Last program version will us print the last cell (i.e., won't indicate S=12, K=12) but an unlikely case. Mote that Pp is now replaced by PI and PZ

(j)ω Ó ***---***---**---| * * * * | C2=1300.6669 (1=188.8990 Ħ F1=0,7500 F2=0,5000 | * * * | a=6. AANA H=12. 8<u>0</u>89 b=2. AHAA H=A. 75AB ーオネチ (J *** * * s)

Appendix E

...

•.

The appendices to Chapter III.

APPENDIX A

An Algorithm, A Computer Code, and A User's Guide, for a Bayesian Binomial Hypothesis Testing Procedure

A.1. INTRODUCTION

In the Bayesian binomial hypothesis testing procedure, we need to find the pair (n_t, x_t^*) such that [see Equations (4) and (5)]:

$$\int_{0}^{1} \sum_{j=0}^{x^{*}} {n \choose j} p_{t}^{j} (1 - p_{t})^{n} t^{-j} g(p_{t}) dp_{t} \leq \alpha$$

and

$$\int_{\Delta}^{1} \sum_{j=0}^{x_{t}^{*}} {n_{t} \choose j} (p_{t} - \Delta)^{j} (1 - p_{t} + \Delta)^{n_{t}-j} g(p_{t}) dp_{t} \ge 1 - \beta,$$

where

$$g(p_t) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} p_t^{\gamma-1} (1 - p_t)^{\delta-1} .$$

The above inequalities can be rewritten as:

$$g_{1}(\mathbf{x}_{t}^{\star},\mathbf{n}_{t}) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \sum_{\mathbf{j}=0}^{\mathbf{x}_{t}^{\star}} {n \choose \mathbf{j}} \frac{\Gamma(\mathbf{j}+\gamma)\Gamma(\mathbf{n}_{t}-\mathbf{j}+\delta)}{\Gamma(\mathbf{n}_{t}+\gamma+\delta)} , \quad (8A)$$

$$g_{2}(\mathbf{x}_{t}^{\star},\mathbf{n}_{t}) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \Delta^{\mathbf{n}_{t}} \sum_{\mathbf{j}=0}^{\mathbf{x}_{t}^{\star}} {n \choose \mathbf{j}} \left[\sum_{\ell=0}^{\mathbf{j}} {j \choose \ell} \Delta^{-\ell}(-1)^{\mathbf{j}-\ell} \right]$$

$$\cdot \left\{ \sum_{m=0}^{\mathbf{n}_{t}-\mathbf{j}} {n \choose m} \Delta^{-m} B(\Delta,1; \ell+\delta, m+\delta) \right\} \geq 1 - \beta , \quad (9A)$$

where

$$B(\Delta,1; r,s) = \int_{\Delta}^{1} p_{t}^{r-1} (1 - p_{t})^{s-1} dp_{t}.$$

A computer code designed to obtain the smallest values of n_t , x* subject to the two inequalities (8A) and (9A), based on an enumeration procedure discussed next, is obtained.

A.2 DESCRIPTION OF THE ENUMERATION PROCEDURE

The enumeration procedure exploits the fact that both $g_1(x_t,n_t)$ and $g_2(x_t,n_t)$ are increasing functions of x_t if n_t is fixed. The procedure starts with some initial value of n_t , say n_t^0 , and finds the largest x_t such that $g_1(x_t,n_t^0) \leq \alpha$. Once such an x_t , say x_t^0 , is found, it is guaranteed that the first inequality will be satisfied for values of x_t smaller than x_t^0 . The procedure then tries to find an x_t smaller than x_t^0 such that $g_2(x_t,n_t) \geq 1 - \beta$. If such an x_t does not exist, the value of n_t is increased by one and the procedure starts all over again. As n_t increases, the procedure finds the smallest values of n_t and x_t satisfying both inequalities. The flow chart for this enumeration procedure is presented in Figure A.1.

A.3 THE COMPUTER CODE

The program requires certain JCL cards and a user input of some parameters.



20

۰.

A.3.1 Input Specifications

The cards should be arranged in accordance with Figure A.2; each card will be explained individually.

Job Card and JCL Cards: The standard job card is used and so are the following JCL cards:

//#EXEC#FORG2

//FORT.SYSINDD

//GO.SYSLIB#DD

//#####DD####DSN=GWU.IMSL.V9.DLOAD,DISP=SHR

//GO.SYSINARADDARARA*

where the character """ indicates a blank space. The first two JCL cards immediately follow the job card. The remaining JCL cards are placed after the program and just before the input information card. The fourth JCL card is needed to use the IMSL subroutines on an IBM machine.

Input Information Card--DEL, SGM, SDEL, ALF, BETA, NT: This card contains sorted input information, DEL, SGM, and SDEL, which are the parameters Δ , γ , and δ in Equations (8A) and (9A); ALF and BETA are the right-hand side parameters α and β in these inequalities. These parameters are specified in format F10.5. The input NT is the initial value of n, selected, and is in I4 format. Usually, this value is one.



Figure A.2. Card deck structure.

A.3.2 Interpretation of Output

The program uses an iterative scheme and evaluates $g_1(x_t, n_t)$ and $g_2(x_t, n_t)$ for different values of x_t add n_t . On the output, the values of $g_1(x_t, n_t)$ and $g_2(x_t, n_t)$ are printed as

> FIRST CONST = SECOND CONST =

for different values of x_t and n_t .

The solution of the problem, that is, the smallest values of x_t and n_t satisfying the inequalities (8A) and (9A), are printed in the

last line of the output as

X = N =

Sample output is presented in Table A.1.

The smallest values of x_t and n_t satisfying the inequalities (8A) and (9A) are X = 10 and N = 15. In this example, the values of the parameters are $\Delta = 0.25$, $\gamma = 106$, $\delta = 19$, $\alpha = 0.10$, and $\beta = 0.25$. The initial value of n_t is one.

The listing of the program is given in Appendix B.

TABLE A.1

Sample Output

		the second s	
		_	
FIRST CONST=	2.2000	$XT = 0 \cdot 0$	NT = 12
FIRST CJAST=	2.0000	XT= 1.0	NT= 12
FIRST COAST=)•))))))	$X\Gamma = 2 \cdot 0$	NT= 12
FIRST CONST=	0.00002	XF= 3.0	NT = 12
FIRST CONST=	0.00019	ΧT= 4.C	NT = 12
FIRST CONST=	0.00135	$X\Gamma = 5.0$	¥T = 12
FIRST CONST=	0.00736	XT = 6.0	N ⁿ = 12
FIRST CONST=	0.03140	Ϋ́Γ= 7.0	NT- 10
FIRST CONTE	3.10523	YF= 8 0	NE 12
SECOND CONST=	1.55776		
FTRSP COUST-	2 202275		
FIRST COVET-		$A = \bigcup_{i=1}^{n} A_{i}$	N2= 13
FINER CONSTR-		$x_{1} = 1 \cdot 0$	11 = 13
FIRST CONSTE	2.0.19.2	y t = 2.0	NT = 13
ELEST CONSTE	3.24921	$\Sigma \Gamma = 3 \cdot C$	<u>NT</u> = 13
*1851 C)437=).)))))	$X \Gamma = 4 \cdot 0$	₹T = 13
FIFST CONST=	7.00041	УТ= 5.0	NT= 13
FIFSI COUST=	2.23245	$\lambda \Gamma = -6 \cdot 0$	VT= 13
FIBSE COAST=	0.21157	$X \Gamma = 7 \cdot 0$	∛] = 13
FIRST CONST=	5.34379	T = - C	$\Lambda T = 13$
FIESE CONST=	1.132-3	VT= 9.0	VT= 13
TECCND TUIST=	3.55554	$r r = B \cdot r$	vr- 13
51-PST 1045T=	1,170		NT- 10
ETEST CONST-		· · · · · · · ·	
	1 1 1 1 1 1		$N_{\rm F} = -1.4$
PIPET CONST-	$\mathbf{J} \bullet \mathbf{J} \mathbf{J} \mathbf{J} \mathbf{J} \mathbf{J} \mathbf{J}$	$\mathbf{x} \mathbf{r} = 2 \cdot 0$	10= 14
		$\mathbf{X} = \mathbf{A} \mathbf{G}$	$N^{T} = 14$
	3.300.2	<u>NT</u> = 4.0	NT = 14
	J.JJ012	XI= 5.0	9 <u>7</u> = 14
FIFSE CDASE=	3.00057	XD= 5.0	NT = 14
EF8SE CONGTE	0.00410	XF= 7.0	YT= 14
FIBSE CONSTE	1717	X 🖞 = 🛛 - C	4T = 14
FIRST CONST=	0.03857	XT= 9.0	NT = 14
FIRST CONST=	7.16215	¥T= 10.0	NT= 14
SECOND CONSTE	0.71435	KT= 9.0	NT= 14
FIPST CONST=	0.00000	XI= 0.0	NT= 15
FIRST CONST=	0.00000	$\chi \Gamma = 1 \cdot 0$	YF= 15
FIRST CUNST=	3.33333	XT = 2.0	VP= 15
FIRST COAST=	0.00000	$\mathbf{XT} = 3 \cdot 2$	NT- 15
FIRST CONFT=	2,20222	$\mathbf{X} \mathbf{T} = \mathbf{u} \mathbf{C}$	N = 15
FIRST COVIT=	1 1121		12 15
			NE= 15
			N = 15
	J.J.2141	$\mathbf{X}\mathbf{F} = -7 \cdot 0$	NT= 15
71858 JJN38=	1.1040	γΓ= 3.0	VI= 15
FIRST CONST =	1.02432	λT= 9.6	NT = 15
FIRST CONSTE	3.757°	X7= 10+0	NT= 15
FIRST COMST=	0.19347	YE= 11.0	SE= 15
STCOND CONSPE	78264	2.T= 10.0	NT= 15
RECORD CONSI=	1.53272	"T= 9.0	NT= 15
7= 10.00	-יי ררו	15	

APPENDIX B

A Listing of the Program for a Bayesian Binomial Hypothesis Testing Procedure

```
LIM=1.5
J=TEST
// EXEC FORX2
//FORT.SYSIN DD *
      IMPLICIT REAL*8 (A-H,O-Z)
      INTEGER IER
      READ (5, 10) DEL, SGM, SDEL, ALF, BETA, NT
   10 FORMAT (5F10.5,14)
      BET=1.0-BETA
      X1=DEL
      X2=1.0
C WE START THE ALGORITHM BY INITIATING XT AS ZERO
      W1=SGM
      W2=SDEL
      A1=W1
      B1=W2
      CALL FACT1 (A1, B1, SON)
      W=SON
   1) XT=0.0
      WNT=NT
      W4=WNT+SDEL
      TA1=SGM
      TB1=W4
      CALL FACT2 (TA1, TB1, TERS)
      PAR=TERS
      CO1=W*PAR
C THIS IS THE VALUE WHEN XT IS ZERO
C NOW WE COMPUTE THE VALUE GI WHEN XT IS OTHER THAN ZERO.
  301 IXT=XT
      T0T=C01
      IF (XT.EQ.0.0) GO TO 1001
      DO 1000 I=1,IXT
      R = I
      P1=W1+R1
      P2=W4-R1
      TA1=P1
      TB1=P2
      CALL FACT2 (TA1, TB1, TERS)
      P3=WNT+1.0
      P4=P3-R1
      P5=R1+1.0
      Z = (DGAMMA(P3)) / ((DGAMMA(P4)) * (DGAMMA(P5)))
      P=TERS
      TOT=TOT+(P*Z*W)
 1000 CONTINUE
 1001 G1=T0T
      WRITE (6,60) G1, XT, NT
   60 FORMAT (5X, 'FIRST CONST=', F10.5, 5X, 'XT=', F5.1, 5X, 'NT=', 14)
C SO WE COMPUTED THE VALUE OF FIRST CONSTRAINT
      IF (G1.GT.ALF) GO TO 333
      IF (XT.EQ.NT) GO TO 380
      XT=XT+1.0
      GO TO 301
  333 XT=XT-1.0
```

.

```
PAGE 2
```

IF (XT.LT.O.O) GO TO 999 C OTHERWISE WE GO AND CALCULATE G2 380 WW=W* (DEL**WNT) C NOW COMPUTE THE VALUE WHEN XT IS ZERO, THAT IS J IS ZERO. C WHEN J IS ZERO L IS ZERO C WHEN J IS' ZERO, M GOES FROM ZERO TO NT AND L IS ALWAYS ZERO IN THIS CASE C FIRST CONSIDER THE CASE WHERE WHEN M IS ZERO A=W1 B=W2 TA1=W1 TB1=W2 CALL FACT2 (TA1, TB1, TERS) CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, IER) Y=TERS VALO= (P2-P1) *Y SUM=VALO C NOW CONSIDER THE CASES WHERE M IS ONE TO NT. DO 1500 M=1.NT A=W1 BM=M BM1=WNT+1.0 BH2=WNT-BM+1.0 BM3=BM+1.0 BMCOM=DGAMMA (BM1) / ((DGAMMA (BM2)) * (DGAMMA (BM3))) BFAC= (DEL** (-BM)) *BMCOM B=W2+BM TA1=W1 TB1=BCALL FACT2 (TA1, TB1, TERS) CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, 1ER) Y=TERS VAL = (P2 - P1) + Y + BFACSUM=SUM+VAL 1500 CONTINUE JXT=XT RJSUM=SUM C IF XT IS ZERO WE HAVE ONLY THE ABOVE TERM IF (XT.EQ.0.0) GO TO 2001 D0 2000 J=1,JXT C THIS IS THE MOST OUTER SUM RJ=J RJ1=WNT+1.0 RJ2=WNT-RJ+1.0 RJ3=RJ+1.0 COMBJ = (DGAMMA(RJ1)) / ((DGAMMA(RJ2)) * (DGAMMA(RJ3)))C NOW L IS FROM ZERO TO J.AGAIN CONSIDER THE CASE WHERE L IS ZERO LP= (~1) **J PL=LP C NOTE WHEN L IS ZERO M GOES FROM ZERO TO NT-J LJL=NT-J 1F (LJL.EQ.0) GO TO 2101 DO 2100 M=1.LJL

•

. •

```
PAGE 3
```

```
RRM=M
      RRM1=WNT-RJ+1.0
      RRM2=WNT-RJ-RRM+1.0
      RRM3=RRM+1.0
      RCOM= (DGAMMA (RRM1)) / ( (DGAMMA (RRM2) ) * (DGAMMA (RRM3) ) )
      FFAC= (DEL** (~RRM)) *RCOM
      A=SGM
      B=RRM+SDEL
      TA1=A
      TB1=B
      CALL FACT2 (TA1, TB1, TERS)
      CALL MDBETA (X1, A, B, P1, IER)
      CALL MDBETA (X2, A, B, P2, IER)
      Y=TERS
      VALM= (P2-P1) *FFAC*Y
      VALO=VALO+VALM
 2100 CONTINUE
 2101 RLSUM=VALO*PL
C THIS IS THE VALUE WHEN L IS ZERO
C NOW WE WANT TO CONSIDER L FROM 1 TO J.THIS IS THE SECOND SUM
      DO 2500 L=1,J
      RL=L
      RL1=RJ-RL+1.0
      RL2=RL+1.0
      COMBL= (DGAMMA (RJ3)) / ( (DGAMMA (RL1)) * (DGAMMA (RL2)))
      LPL = (-1) \star \star (J-L)
      FLP=LPL
      POWER=DEL** (-RL)
      FACL=FLP*COMBL*POWER
C NOW SHOULD CONSIDER M LOOP AGAIN.NOW M IS FROM ZERO TO NT-J FOR GIVEN L
C START WITH MIS ZERO
      A=RL+SGM
      B=SDEL
      CALL MDBETA (X1, A, B, P1, IER)
      CALL MDBETA (X2, A, B, P2, IER)
      TAI=A
      TB1=B
      CALL FACT2 (TA1, TB1, TERS)
      Y=TERS
      VAL = (P2 - P1) \star Y
      RMSUM=VAL
      LL=NT-J
       IF (LL.EQ.0' GO TO 3001
      DO 3000 M=1,LL
      RM=M
      RM1=WNT-RJ+1.0
      RM2=WNT-RJ-RM+1.0
      RM3=RM+1.0
      COMBM= (DGAMMA (RM1)) / ( (DGAMMA (RM2) ) * (DGAMMA (RM3) ) )
      FACM=(DEL \star \star (-RM)) \star (COMBM)
      A=RL+SGM
      B=RM+SDEL
      CALL MDBETA (X1, A, B, P1, IER)
      CALL MOBETA (X2, A, B, P2, IER)
```

```
PAGE 4
```

TA1=A TB1=B CALL FACT2 (TA1, TB1, TERS) Y=TERS VAL = (P2 - P1) * FACM * YRMSUM=RMSUM+VAL 3000 CONTINUE 3001 RRSUM=RMSUM C THE MOST INNER LOOP IS FINISHED. RLSUM= (FACL*RRSUM) +RLSUM C THIS IS THE SUM FOR L LOOP 2500 CONTINUE C L LOOP IS FINISHED C NOW FINISH J LOOP. THE MOST OUTER LOOP. RJSUM= (COMBJ*RLSUM) +RJSUM 2000 CONTINUE C SO WE EVALUATED G2. 2001 G2=RJSUM*WW WRITE (6,61) G2, XT, NT 61 FORMAT (5X, 'SECOND CONST=', F10.5, 5X, 'XT=', F5.1, 5X, 'NT=', 14) IF (G2.LT.BET) GO TO 999 777 IF (XT.LT.1.0) GO TO 888 XT=XT-1.0 C CHECK G2 AGAIN. WW=W* (DEL**WNT) C NOW COMPUTE THE VALUE WHEN XT IS ZERO, THAT IS J IS ZERO. C WHEN J IS ZERO L IS ZERO. C WHEN J IS ZERO, M GOES FROM ZERO TO NT AND L IS ALWAYS ZERO IN THIS CASE C FIRST CONSIDER THE CASE WHERE WHEN M IS ZERO A=W) B=W2 CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, IER) TA1=A TB1=B CALL FACT2 (TA1, TB1, TERS) Y=TERS VALO=(P2-P1)*YSUM=VALO C NOW CONSIDER THE CASES WHERE M IS ONE TO NT. DO 1501 M=1,NT A=W1 BM=M BM1=WNT+1.0 BM2=WNT-BM+1.0 BM3=BM+1.0 BMCOM=DGAMMA (BM1) / ((DGAMMA (BM2)) * (DGAMMA (BM3))) BFAC= (DEL** (-BM)) *BMCOM B=W2+BM TA1=A TB1=B CALL FACT2 (TA1, TB1, TERS) CALL MOBETA (X1, A, B, P1, IER) CALL MOBETA (X2, A, B, P2, IER)

<u>,</u>

PAGE 5

Y=TERS VAL= (P2-P1) *Y*BFAC SUM=SUM+VAL 1501 CONTINUE JXT=XT RJSUM=SUM C IF XT IS ZERO WE HAVE ONLY THE ABOVE TERM IF (XT.EQ.0.0) GO TO 2011 DO 5000 J=1, JXT C THIS IS THE MOST OUTER SUM RJ≠J RJ1=WNT+1.0 RJ2=WNT-RJ+1.0 RJ3=RJ+1.0 COMBJ= (DGAMMA (RJ1)) / ((DGAMMA (RJ2)) * (DGAMMA (RJ3))) C NOW L IS FROM ZERO TO J.AGAIN CONSIDER THE CASE WHERE L IS ZERO LP= (-1) **J PL=LP C NOTE WHEN L IS ZERO M GOES FROM ZERO TO NT-J LJL=NT-J IF (LJL.EQ.0) GO TO 2102 DO 2105 M=1,LJL RRM=M RRM1=WNT-RJ+1.0 RRM2=WNT-RJ-RRM+1.0 RRM3=RRM+1.0 RCOM= (DGAMMA (RRM1)) / ((DGAMMA (RRM2)) * (DGAMMA (RRM3))) FFAC= (DEL** (-RRM)) *RCOM A=SGM B=RRM+SDEL CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, IER) TA = ATB1=B CALL FACT2 (TA1, TB1, TERS) Y=TERS VALM= (P2-P1) *FFAC*Y VALO=VALO+VALM 2105 CONTINUE 2102 RLSUM=VALO*PL C THIS IS THE VALUE WHEN L IS ZERO C NOW WANT TO CONSIDER L FROM 1 TO J. THIS IS THE SECOND SUM DO 2501 L=1,J RL=L RL1=RJ-RL+1.0 RL2=RL+1.0 COMBL = (DGAMMA(RJ3)) / ((DGAMMA(RL1)) * (DGAMMA(RL2)))LPL = (-1) ** (J-L)FLP=LPL POWER=DEL** (-RL) FACL=FLP*COMBL*POWER C NOW SHOULD CONSIDER M LOOP AGAIN.NOW M IS FROM ZERO TO NT-J FOR GIVEN L C START WITH MIS ZERO. A=RL+SGM

.

```
PAGE 6
```

B=SD£L CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, IER) TA !=A TBi=B CALL FACT2 (TA1, TB1, TERS) Y=TERS VAL= (P2-P1) *Y RMSUM=VAL LL=NT-J IF (LL.EQ.0) GO TO 4001 DO 4000 M=1,LL RM=M RM1=WNT-RJ+1.0 RM2=WNT-RJ-RM+1.0 RM3=RM+1.0 COMBM= (DGAMMA (RM1)) / ((DGAMMA (RM2)) * (DGAMMA (RM3))) $FACM=(DEL \pi \pi (-RM)) \pi (COMBM)$ A=RL+SGM B=RM+SDEL CALL MDBETA (X1, A, B, P1, IER) CALL MDBETA (X2, A, B, P2, IER) TA1=A TB1=BCALL FACT2 (TA1, TB1, TERS) Y=TERS VAL = (P2 - P1) * FACM * YRMSUM=RMSUM+VAL 4000 CONTINUE 4001 RRSUM=RMSUM C THE MOST INNER LOOP IS FINISHED. RLSUM= (FACL*RRSUM) +RLSUM C THIS IS THE SUM FOR L LOOP 2501 CONTINUE C L LOOP IS FINISHED. C NOW FINISH J LOOP. THE MOST OUTER LOOP. RJSUM= (COMBJ*RLSUM) +RJSUM 5000 CONTINUE C SO WE EVALUATED G2. 2011 G2=RJSUM*WW WRITE (6,62) G2, XT, NT 62 FORMAT (5X, 'SECOND CONST='F10.5, 5X, 'XT=', F5.1, 5X, 'NT=', 14) C CHECK G2 NOW IF (G2.GE.BET) GO TO 777 XT=XT+1.0 GO TO 888 999 NT=NT+1 GO TO 11 888 WRITE (6,555) XT.NT 555 FORMAT (10X, 'X=', F10.5, 5X, 'N=', 14) STOP END SUBROUTINE FACT1 (A1, B1, SON) IMPLICIT REAL*8 (A-H, 0-Z)

•••

C=A1+B1 IF (A'I.LE.57.0.AND.C.LE.57.0) GO TO 41 C1 = C - 1.0A2=A1-1.0 B2=B1-1.0 C2=A2+B2 1B=A2+1.0 1C=C2 PAY=C1 DO 42 1=18,1C Z1=1 PAY=PAY+Z1 42 CONTINUE PAYDA=1.0 JA=B2 DO 43 J=1, JA VJ=J PAYDA=PAYDA*VJ 43 CONTINUE SON=PAY/PAYDA GO TO 45 41 SON=DGAMMA(C)/((DGAMMA(A1)) + (DGAMMA(B1))) 45 CONTINUE RETURN END SUBROUTINE FACT2 (TA1, TB1, TERS) IMPLICIT REAL*8 (A-H, 0-Z) C=TA1+TB1 IF (TA1.LE.57.0.AND.C.LE.57.0) GO TO 71 C1 = C - 1.0A2=TA1-1.0 B2=TB1-1.0 C2=A2+B2 1B=A2+1.0 1C=C2PAY=C1 DO 72 1=18,1C Z1=1 PAY=PAY*ZI 72 CONTINUE PAYDA=1.0 JA=B2 DO 73 J=1, JA VJ=J PAYDA=PAYDA*VJ 73 CONTINUE TERS=PAYDA/PAY GO TO 75 71 TERS=((DGAMMA (TA1)) * (DGAMMA (TB1)))/(DGAMMA (C)) 75 CONTINUE RETURN END //GO.SYSLIB DD DD DSN=GWU.IMSL.V9.DLOAD,DISP=SHR 11

.

-

//GD.SYSIN DD * 0.25000 113.00000 20.00000 0.10100 0.25000 1 //

•

-- 2

• • •

PAGE 8

APPENDIX C

Illustrative Calculation of Expected Sample Sizes for Curtailed Sequential Sampling

C.1. THE CASE OF TESTING ONE ITEM AT A TIME

t

We illustrate this for Stage 0. Here $x_j^* = 5$, $n_t = 17$. We must have either 6 successes to accept, or 12 failures to reject

$$P[n_{t}=6|p_{t}] = {5 \choose 5} p_{t}^{6} = 0.015625$$

$$P[n_{t}=7|p_{t}] = {6 \choose 5} p_{t}^{6}(1-p_{t}) = 0.046875$$

$$P[n_{t}=8|p_{t}] = {7 \choose 5} p_{t}^{6}(1-p_{t})^{2} = 0.0820312$$

$$P[n_{t}=9|p_{t}] = {8 \choose 5} p_{t}^{6}(1-p_{t})^{3} = 0.109375$$

$$P[n_{t}=10|p_{t}] = {9 \choose 5} p_{t}^{6}(1-p_{t})^{4} = 0.1230469$$

$$P[n_{t}=11|p_{t}] = {10 \choose 5} p_{t}^{6}(1-p_{t})^{5} = 0.1230469$$

$$P[n_{t}=12|p_{t}] = {11 \choose 5} p_{t}^{6}(1-p_{t})^{6} + {11 \choose 11} (1-p_{t})^{12} = 0.1130371$$

$$P[n_{t}=13|p_{t}] = {12 \choose 5} p_{t}^{6}(1-p_{t})^{7} + {12 \choose 11} p_{t}(1-p_{t})^{12} = 0.0968018$$

$$P[n_{t}=14|p_{t}] = {13 \choose 5} p_{t}^{6}(1-p_{t})^{8} + {13 \choose 11} p_{t}^{2}(1-p_{t})^{12} = 0.083313$$

$$P[n_{t}=15|p_{t}] = {\binom{14}{5}} p_{t}^{6}(1-p_{t})^{9} + {\binom{14}{11}} p_{t}^{3}(1-p_{t})^{12} = 0.0722046$$

$$P[n_{t}=16|p_{t}] = {\binom{15}{5}} p_{t}^{6}(1-p_{t})^{10} + {\binom{15}{11}} p_{t}^{4}(1-p_{t})^{12} = 0.0666504$$

$$P[n_{t}=17|p_{t}] = {\binom{16}{5}} p_{t}^{6}(1-p_{t})^{11} + {\binom{16}{11}} p_{t}^{5}(1-p_{t})^{12} = 0.0666504$$

To obtain $P[n_t=j]$, j = 6,7,...,17, we average out the above by using $g(p_t|\cdot)$. At Stage 0, $\gamma = 1$, $\delta = 1$.

$$\begin{split} p[n_t=6] &= \frac{\Gamma(\gamma+6)\Gamma(\delta)}{\Gamma(\gamma+\delta+6)} = 0.1428571 \\ p[n_t=7] &= 6 \frac{\Gamma(\gamma+6)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+7)} = 0.1071429 \\ p[n_t=8] &= 21 \frac{\Gamma(\gamma+6)\Gamma(\delta+2)}{\Gamma(\gamma+\delta+8)} = 0.083333 \\ p[n_t=9] &= 56 \frac{\Gamma(\gamma+6)\Gamma(\delta+3)}{\Gamma(\gamma+\delta+9)} = 0.0666667 \\ p[n_t=10] &= 126 \frac{\Gamma(\gamma+6)\Gamma(\delta+4)}{\Gamma(\gamma+\delta+10)} = 0.0545455 \\ p[n_t=11] &= 252 \frac{\Gamma(\gamma+6)\Gamma(\delta+5)}{\Gamma(\gamma+\delta+11)} = 0.0454545 \\ p[n_t=12] &= 402 \frac{\Gamma(\gamma+6)\Gamma(\delta+6)}{\Gamma(\gamma+\delta+12)} + \frac{\Gamma(\gamma)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+12)} = 0.1103896 \\ p[n_t=13] &= 792 \frac{\Gamma(\gamma+6)\Gamma(\delta+7)}{\Gamma(\gamma+\delta+13)} + 12 \frac{\Gamma(\gamma+1)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+13)} = 0.0989011 \\ p[n_t=14] &= 1287 \frac{\Gamma(\gamma+6)\Gamma(\delta+8)}{\Gamma(\gamma+\delta+14)} + 78 \frac{\Gamma(\gamma+2)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+14)} = 0.0857143 \\ p[n_t=15] &= 2002 \frac{\Gamma(\gamma+6)\Gamma(\delta+9)}{\Gamma(\gamma+\delta+15)} + 364 \frac{\Gamma(\gamma+3)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+16)} = 0.066176 \\ p[n_t=17] &= 4368 \frac{\Gamma(\gamma+6)\Gamma(\delta+11)}{\Gamma(\gamma+\delta+17)} + 4368 \frac{\Gamma(\gamma+5)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+17)} = 0.0588235 \\ E[n_t] &= 10.91 . \end{split}$$

:

C.2. THE CASE OF TESTING IN BATCHES OF SIZE 3

Stage 0

ĩ

$$p_t = 0.5 (1-p_t) = 0.5$$

 $x_j^* = 5 \quad n_t = 17$

We must have either 6 successes to accept or 12 failures to reject.
Thus,
$$n_t c\{6,9,12,15,18\}$$

 $p[n_t = 6 | p_t] = {6 \choose 6} p_t^6 = 0.015625$
 $p[n_t = 9 | p_t] = {6 \choose 5} p_t^6 (1 - p_t) + {7 \choose 5} p_t^6 (1 - p_t)^2 + {8 \choose 5} p_t^6 (1 - p_t)^3 = 0.2382812$
 $p[n_t = 12 | p_t] = {9 \choose 5} p_t^6 (1 - p_t)^4 + {10 \choose 5} p_t^6 (1 - p_t)^5 + {11 \choose 5} p_t^6 (1 - p_t)^6$
 $+ {12 \choose 12} (1 - p_t)^{12} = 0.3444824$
 $p[n_t = 15 | p_t] = {12 \choose 5} p_t^6 (1 - p_t)^7 + {13 \choose 5} p_t^6 (1 - p_t)^8 + {14 \choose 5} p_t^6 (1 - p_t)^9$
 $+ {12 \choose 11} p_t (1 - p_t)^{12} + {13 \choose 11} p_t^2 (1 - p_t)^{12} + {14 \choose 11} p_t^3 (1 - p_t)^{12} = 0.2536621$
 $p[n_t = 18 | p_t] = {15 \choose 15} p_t^5 (1 - p_t)^{10} + {15 \choose 4} p_t^4 (1 - p_t)^{11} = 0.1333008$

Stage 1

$$p_t = 0.875 (1-p_t) = 0.125$$

 $x_t^* = 9 \quad n_t = 13$

We must have either 10 successes to accept or 4 failures to reject. Thus, $n_t \in \{6,9,12,15\}$

$$p[n_{t}=6|p_{t}] = {\binom{6}{4}} p_{t}^{2}(1-p_{t})^{4} + {\binom{6}{5}} p_{t}(1-p_{t})^{5} + {\binom{6}{6}} (1-p_{t})^{6} = 0.0029678$$
$$p[n_{t}=9|p_{t}] = {\binom{6}{3}} p_{t}^{3}(1-p_{t})^{4} + {\binom{7}{3}} p_{t}^{4}(1-p_{t})^{4} + {\binom{8}{3}} p_{t}^{5}(1-p_{t})^{4} = 0.0140249$$

$$p[n_{t}=12|p_{t}] = {\binom{9}{3}} p_{t}^{6}(1-p_{t})^{4} + {\binom{10}{3}} p_{t}^{7}(1-p_{t})^{4} + {\binom{11}{3}} p_{t}^{8}(1-p_{t})^{4} + {\binom{9}{9}} p_{t}^{10}$$

$$+ {\binom{10}{9}} p_{t}^{10}(1-p_{t}) + {\binom{11}{9}} p_{t}^{10}(1-p_{t})^{2} = 0.852551$$

$$p[n_{t}=15|p_{t}] = {\binom{12}{3}} p_{t}^{9}(1-p_{t})^{3} = 0.1291889$$

Stage 2

•

$$p_t = 0.9 (1-p_t) = 0.1$$

 $x_t^* = 8 n_t = 11$

We must have either 9 successes to accept or 3 failures to reject. Thus, $n_t \epsilon \{3, 6, 9, 12\}$

$$p[n_{t}=3|p_{t}] = {3 \choose 3} (1-p_{t})^{3} = 0.001$$

$$p[n_{t}=6|p_{t}] = {3 \choose 2} p_{t}(1-p_{t})^{3} + {4 \choose 2} p_{t}^{2}(1-p_{t})^{3} + {5 \choose 2} p_{t}^{3}(1-p_{t})^{3} = 0.01485$$

$$p[n_{t}=9|p_{t}] = {6 \choose 2} p_{t}^{4}(1-p_{t})^{3} + {7 \choose 2} p_{t}^{5}(1-p_{t})^{3} + {8 \choose 2} p_{t}^{6}(1-p_{t})^{3} + {9 \choose 0} p_{t}^{9} = 0.4245426$$

$$p[n_{t}=12|p_{t}] = {9 \choose 2} p_{t}^{7}(1-p_{t})^{2} + {9 \choose 1} p_{t}^{8}(1-p_{t}) = 0.5596074$$

•••

Stage 3

 $p_t = 0.906$ $1-p_t = 0.094$ $x_t^* = 8$ $n_t = 11$

The same enumeration as in Stage 2.

Stage 4

$$p_t = 0.909 \quad 1-p_t = 0.091$$

 $x_t^* = 8 \quad n_t = 11$

The same enumeration as in Stage 2.

Stage 5

5 F

 $p_t = 0.875 \quad 1 - p_t = 0.125$ $x_t^* = 9 \quad n_t = 13$

The same enumeration as in Stage 1.

Stage 6

 $p_t = 0.853 \quad 1 - p_t = 0.147$ $x_t^* = 8 \qquad n_t = 12$

We must have either 4 failures to reject or 9 successes to accept. Thus, $n_{t} \in \{6,9,12\}$

$$p[n_{t}=6|p_{t}] = {\binom{6}{4}} p_{t}^{2}(1-p_{t})^{4} + {\binom{6}{5}} p_{t}(1-p_{t})^{5} + {\binom{6}{6}} (1-p_{t})^{6} = 0.0054577$$

$$p[n_{t}=9|p_{t}] = {\binom{6}{3}} p_{t}^{3}(1-p_{t})^{4} + {\binom{7}{3}} p_{t}^{4}(1-p_{t})^{4} + {\binom{8}{3}} p_{t}^{5}(1-p_{t})^{4} + {\binom{9}{0}} p_{t}^{9} = 0.2653362$$

$$p[n_{t}=12|p_{t}] = {\binom{9}{3}} p_{t}^{6}(1-p_{t})^{3} + {\binom{9}{2}} p_{t}^{7}(1-p_{t})^{2} + {\binom{9}{1}} p_{t}^{8}(1-p_{t}) = 0.7292061$$

Stage 7

$$p_t = 0.825 (1-p_t) = 0.175$$

 $x_j^* = 9 \quad n_t = 14$

We must have either 10 successes to accept or 5 failures to reject. Thus, $n_{t} \in \{6,9,12,15\}$

$$p[n_{t}=6|p_{t}] = {6 \choose 5} p_{t}(1-p_{t})^{5} + {6 \choose 6} (1-p_{t})^{6} = 0.0008412$$

$$p[n_{t}=9|p_{t}] = {6 \choose 4} p_{t}^{2}(1-p_{t})^{5} + {7 \choose 4} p_{t}^{3}(1-p_{t})^{5} + {8 \choose 4} p_{t}^{4}(1-p_{t})^{5} = 0.0102237$$

38

$$p[n_{t}=12|p_{t}] = \begin{pmatrix} 9\\ 4 \end{pmatrix} p_{t}^{5}(1-p_{t})^{5} + \begin{pmatrix} 10\\ 4 \end{pmatrix} p_{t}^{6}(1-p_{t})^{5} + \begin{pmatrix} 11\\ 4 \end{pmatrix} p_{t}^{7}(1-p_{t})^{5} + \begin{pmatrix} 9\\ 9 \end{pmatrix} p_{t}^{10}(1-p_{t})^{5} + \begin{pmatrix} 11\\ 9 \end{pmatrix} p_{t}^{10}(1-p_{t})^{2} = 0.6805573$$
$$p[n_{t}=15|p_{t}] = \begin{pmatrix} 12\\ 4 \end{pmatrix} p_{t}^{8}(1-p_{t})^{4} + \begin{pmatrix} 12\\ 9 \end{pmatrix} p_{t}^{9}(1-p_{t})^{3} = 0.3083778$$

Stage 8

 $p_t = 0.833 (1-p_t) = 0.167$ $x_t^* = 9 \qquad n_t = 14$

The same enumeration as in Stage 7.

Stage 9

 $p_t = 0.820$ (1- p_t) = 0.180 $x_t^* = 9$ $n_t = 14$

The same enumeration as in Stage 7.

Stage 10

$$p_t = 0.837 (1-p_t) = 0.163$$

 $x_t^* = 9 \quad n_t = 14$

The same enumeration as in Stage 7.

Stage 11

$$p_t = 0.841 (1-p_t) = 0.159$$

 $x_t^* = 10 \quad n_t = 15$

We must have either 11 successes to accept or 5 failures to reject. Thus, $n_t \in \{6,9,12,15\}$

$$p[n_{t}=6|p_{t}] = {6 \choose 5} p_{t}(1-p_{t})^{5} + {6 \choose 6} (1-p_{t})^{6} = 0.0005289$$

$$p[n_{t}=9|p_{t}] = {6 \choose 4} p_{t}^{2}(1-p_{t})^{5} + {7 \choose 4} p_{t}^{3}(1-p_{t})^{5} + {8 \choose 4} p_{t}^{4}(1-p_{t})^{5} = 0.0067523$$

$$p[n_{t}=12|p_{t}] = {9 \choose 4} p_{t}^{5}(1-p_{t})^{5} + {10 \choose 4} p_{t}^{6}(1-p_{t})^{5} + {11 \choose 4} p_{t}^{7}(1-p_{t})^{5} + {10 \choose 10} p_{t}^{11}$$

$$+ {11 \choose 10} p_{t}^{11}(1-p_{t}) = 0.4321114$$

$$p[n_{t}=15|p_{t}] = {12 \choose 4} p_{t}^{8}(1-p_{t})^{4} + {12 \choose 9} p_{t}^{9}(1-p_{t})^{3} + {12 \choose 10} p_{t}^{10} (1-p_{t})^{2} = 0.5606073$$
Stage 12

$$p_t = 0.836 (1-p_t) = 0.164$$

 $x_t^* = 9 n_t = 14$

The same enumeration as in Stage 7.

Stage 13

$$p_t = 0.848 (1-p_t) = 0.152$$

 $x_t^* = 8 \quad n_t = 12$

The same enumeration as in Stage 6.

Stage 14

$$p_t = 0.850 (1-p_t) = 0.150$$

 $x_t^* = 8 \qquad n_t = 12$

The same enumeration as in Stage 6.

To obtain the $E(n_t)$, we average out the above by using $g(p_t|\cdot)$. We illustrate this for Stage 0.

$$\begin{split} p[n_t=6] &= \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(\gamma+\delta)\Gamma(\delta)}{\Gamma(\gamma+\delta+6)} = 0.1428571 \\ p[n_t=9] &= \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[6 \frac{\Gamma(\gamma+6)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+7)} + 21 \frac{\Gamma(\gamma+6)\Gamma(\delta+2)}{\Gamma(\gamma+\delta+8)} + 56 \frac{\Gamma(\gamma+6)\Gamma(\delta+3)}{\Gamma(\gamma+\delta+0)} \right] = 0.2571429 \\ p[n_t=12] &= \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[126 \frac{\Gamma(\gamma+6)\Gamma(\delta+4)}{\Gamma(\gamma+\delta+10)} + 252 \frac{\Gamma(\gamma+6)\Gamma(\delta+5)}{\Gamma(\gamma+\delta+11)} + 402 \frac{\Gamma(\gamma+6)\Gamma(\delta+6)}{\Gamma(\gamma+\delta+12)} \right] \\ &+ \frac{\Gamma(\gamma)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+12)} \right] = 0.2103897 \\ p[n_t=15] &= \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[792 \frac{\Gamma(\gamma+6)\Gamma(\gamma+7)}{\Gamma(\gamma+\delta+13)} + 1287 \frac{\Gamma(\gamma+6)\Gamma(\delta+8)}{\Gamma(\gamma+\delta+14)} + 2002 \frac{\Gamma(\gamma+6)\Gamma(\delta+9)}{\Gamma(\gamma+\delta+15)} \right] \\ &+ 12 \frac{\Gamma(\gamma+1)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+13)} + 78 \frac{\Gamma(\gamma+2)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+14)} + 364 \frac{\Gamma(\gamma+3)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+15)} \right] = 0.2596154 \\ p[n_t=18] &= \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[3003 \frac{\Gamma(\gamma+5)\Gamma(\delta+10)}{\Gamma(\gamma+\delta+15)} + 1365 \frac{\Gamma(\gamma+4)\Gamma(\delta+11)}{\Gamma(\gamma+\delta+15)} \right] = 0.125 \\ E[n_t] &= 11.84 . \end{split}$$

Similarly, we can obtain $E[n_t]$ for other stages.

REFERENCES

BOX, G. E. P. (1982). An apology ecumenism in statistics, MRC Technical Report #2408.

ACKNOWLEDGMENTS

The algorithm of Appendix A is due to Mr. Jerzy Kyparisis, and the computer code was developed by Mr. Refik Soyer. We have benefited from discussions with Professors Box, Draper, Harris, Johnson, Leonard and Woodroofe. The problem was brought to our attention by Dr. Willard of the Office of the Deputy Undersecretary of the Army; his comments have helped us develop the methodology proposed. Comments by Dr. Moore of BRL are also appreciated.