SEQUENTIAL SIGNIFICANCE TESTING
AND ESTIMATION

By

MALCOLM ROSS HEYWORTH

Bachelor of Science
University of Auckland
Auckland, New Zealand
1969

Master of Science
University of Auckland
Auckland, New Zealand
1970

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1974

SEQUENTIAL SIGNIFICANCE TESTING

AND ESTIMATION

Thesis Approved:

J. Leroy Folks
Thesis Adviser

P. Larry Claypool

William J. Louis

I. I. Kotlarski

N N Durham
Dean of the Graduate College

PREFACE AND ACKNOWLEDGEMENTS

I wish to acknowledge my advisor, Dr. Leroy Folks, for tolerating my personal assaults on "academic freedom", whatever it is. I also express my gratitude to Drs. Larry Claypool, W. J. Leivo and I. I. Kotlarski for being on my Advisory Committee.

To appropriate faculty members and students, past and present, I extend a sincere thank you for my not so sober moments in Stillwater, Oklahoma 74074. You have been and will continue to be a vital part of my constitutional salvation.

To people in a faraway country I can but say pathetically little, as usual. Hang in there and keep faith in me. The hometown kid may have become insatiably cynical but he hasn't fully retired the hope of doing something significantly useful for love, peace and freedom.

... long years -
Long, though not very many, ...
some suffering and some tears
Have left us nearly where we had begun:
Yet not in vain our mortal race hath run,
We have had our reward - and it is here;
That we can yet feel gladden'd by the sun,
And reap from earth, sea, joy almost as dear
As if there were no man to trouble what is clear.

- Byron, Child Harold's Pilgrimage,
Canto IV, Stanza CLXXVI, c 1818.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

X, Y, Θ, etc.	random variables
x, y, θ, etc.	values of (i.e. observations on) random variables X, Y, etc., respectively
ln	natural logarithmic transformation, base e
$\equiv$	denote(s), or denoting, or "is identical to" (almost everywhere)
cdf	cumulative distribution function
df	degrees of freedom
i.i.d.	independent and identically distributed
r.v.	random variable
r.v's	random variables
SPRT	Sequential Probability Ratio Test
Φ	cumulative distribution function of the standardized normal distribution
f	probability density function (pdf)

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

When a statistically designed experiment is run to test for "significant differences among treatments", the statistical analysis yields a numerical observation on a test statistic whose distribution (under the null hypothesis of no differences among treatment effects) is known. This numerical value may then be transformed (using the known distribution of the test statistic) into an observed level of significance (of the test statistic under the null hypothesis) and this observed significance level may, under the null hypothesis, be interpreted as a random observation on a random variable which is uniformly distributed on the interval (0, 1), assuming the test statistic is of the continuous type. This then is a measure of the consistency or inconsistency of the observed experimental data with the null hypothesis being tested.

If an experiment is repeated and the results of these repetitions can be treated as independent of one another, a naturally arising question is "How can the experimental data be combined to give an overall set of experimental data so that a meaningful overall analysis can be run on the combined data?" If, for example, two agronomists (sceptical of each other's abilities) run identical completely randomized experiments in neighbouring plots (each experimenter doing an individual randomization, of course), each can analyze his data separately or their data can be easily combined and a meaningful analysis run on this combined data. However, if three experimenters (each oblivious of the other's work) run experiments measuring the one "quantity" (say the differences among a standard treatment at present in use, a new treatment and a "control", i.e., no treatment), one experimenter using a completely randomized design in Fort Collins, Colorado, another using a randomized block design in Ames, Iowa and the third a Latin square in Stillwater, Oklahoma, and their data cannot be easily combined by any known technique to yield a "useful" test statistic, then how can their separate results be combined to yield a meaningful overall result?

Fisher (22, Section 21.1, pages 99-101) suggested the following method. Let $u_1, u_2, \ldots, u_n$ be the observed significance levels of n independent test statistics; then (under the combination of all n null hypotheses) $-2 \ln (\prod_{i=1}^{n} u_i)$ is an observation on a chi-squared random variable with $2n$ degrees of freedom, so <u>an</u> overall significance level for all individual experimental results combined can be determined.

Since the natural logarithmic function ln is one-to-one, Fisher's method is equivalent to multiplying the individual significance levels and determining the significance level of this product. It is easily shown (by induction, for example) that the density of this product random variable T (under the combination of all n null hypotheses) is given by

$$f_n(t) = \begin{cases} \frac{1}{(n-1)!} (-\ln t)^{n-1}, & 0 < t < 1, \\ 0 & \text{otherwise}, \end{cases}$$

$$n = 1, \ldots$$

(so $-2 \ln T \sim \chi^2(2n)$, $n = 1, \ldots$). Thus $x \to -2 \ln x$, $x \varepsilon (0, 1)$, may be regarded as "Fisher's transformation."

Again, Fisher's method is equivalent to transforming each observed significance level into an observation on an exponentially distributed random variable - with common parameter - by a common logarithmic transformation, then summing these observations and determining the significance level of this sum, for if $U_1, \ldots, U_n$ are independent and identically distributed (i.i.d.) random variables with uniform distribution on $(0, 1)$ then $\forall\ \lambda > 0$, $Y_i \equiv -\lambda \ln U_i$ has density given by

$$f(y) = \begin{cases} \frac{1}{\lambda} e^{-\frac{y}{\lambda}}, & y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$i = 1, \ldots, n,$$

and since then $\frac{2}{\lambda} Y_i \sim \text{i.i.d.}\ \chi^2(2)$,

$$\frac{2}{\lambda} \sum_{i=1}^{n} Y_i \sim \chi^2(2n), \quad n = 1, \ldots .$$

Fisher's method has the disadvantage that it does not allow for the significance levels to be weighted. If, for example, u_1 is the significance level of an observation on a chi-squared random variable with one degree of freedom whereas u_2 is the significance level of an observation on a chi-squared variate with one hundred degrees of freedom it seems rational and reasonable to give u_2 one hundred times the weight of u_1, yet Fisher's method does not do this.

According to van Zwet and Oosterhoff (48), Lancaster (32) has given a method of weighting significance levels. I. J. Good (24) and Zelen and Joel (54) have given restricted methods of doing likewise. Good considered the distribution of the variate

$$Q \equiv P_1^{\lambda_1} P_2^{\lambda_2} \cdots P_n^{\lambda_n}, \text{ where } P_1, \ldots, P_n \sim \text{i.i.d. } U(0, 1)$$

and $\lambda_1, \ldots, \lambda_n$ are unequal positive weights,

and showed that $\forall\, q \in [0, 1]$,

$$P(Q < q) = \sum_{k=1}^{n} \Lambda_k\, q^{\frac{1}{\lambda_k}}, \text{ where } \Lambda_1, \ldots, \Lambda_n \text{ are} \tag{1}$$

constants defined by the partial fraction expansion

$$\prod_{k=1}^{n} \frac{1}{1-i\lambda_k t} \equiv \sum_{k=1}^{n} \frac{\Lambda_k}{1-i\lambda_k t} .$$

Property 1: The weights need be known only to within an arbitrary factor since for $\mu_k = \lambda\,\lambda_k$, $k = 1, \ldots, n$, for some $\lambda > 0$,

$$P\Big(\prod_{k=1}^{n} P_k^{\mu_k} < r\Big) = P\Big(\big(\prod_{k=1}^{n} P_k^{\lambda_k}\big)^{\lambda} < r\Big)$$

$$= P\big(Q < r^{\frac{1}{\lambda}}\big)$$

$$= \sum_{k=1}^{n} M_k r^{\frac{1}{\mu_k}}, \text{ where}$$

$$\prod_{k=1}^{n} \frac{1}{1-i\mu_k t} \equiv \sum_{k=1}^{n} \frac{M_k}{1-i\mu_k t} ,$$

so

$$\prod_{k=1}^{n} \frac{1}{1-i\lambda_k u} \equiv \sum_{k=1}^{n} \frac{M_k}{1-i\lambda_k u} , \quad u \equiv \lambda t$$

$$\equiv \sum_{k=1}^{n} \frac{\Lambda_k}{1-i\lambda_k u} ,$$

i.e. $M_k = \Lambda_k$, $k = 1, \ldots, n$.

Thus if, for example, two significance levels are available from chi-squared variates, one with one degree of freedom and the other with two degrees of freedom, then Good's formula (with weights proportional to the number of degrees of freedom of the chi-squared variates underlying the respective significance levels) yields an overall significance level which would be equal to the significance level calculated from the same formula if the given significance levels were obtained from chi-squared variates, one with fifty degrees of freedom and the other with one hundred degrees of freedom.

Modifications of Fisher's method to adapt it to the case where the underlying distribution is discrete have been proposed by Wallis (51), Lancaster (33) and E. S. Pearson (44). Kincaid (31) has written an excellent article clarifying the relationship among these methods. Lancaster suggests that in many cases the observed significance level may be replaced in "Fisher's transformation" with the average of the observed significance level and the next lower level attainable (the lowest level being defined as zero).

The references given so far all have an outstanding singularity of purpose: all deal with a random sample of significance levels of fixed size - none deals with a sequential procedure.

The Problems

The Sequential Probability Ratio Test (SPRT) of Wald (49) is of the following form:

To test the simple hypothesis H_0: $\theta = \theta_0$ against the simple alternative H_1: $\theta = \theta_1$ $(\neq \theta_0)$ calculate the likelihood ratio

$\frac{p_{1m}}{p_{0m}} \equiv \lambda_m$ after the m^{th} random observation has been taken $(m = 1, \ldots)$ and either

(i) accept H_0 if $\lambda_m < \frac{\beta}{1 - \alpha}$, or

(ii) accept H_1 if $\lambda_m > \frac{1 - \beta}{\alpha}$, or

(iii) if $\frac{\beta}{1 - \alpha} < \lambda_m < \frac{1 - \beta}{\alpha}$ then take another observation.

Here p_{im} is the likelihood under H_i, $i = 0, 1$, and α and β are the desired overall probabilities of Types I and II errors, respectively. The SPRT boundaries $\frac{\beta}{1 - \alpha}$ and $\frac{1 - \beta}{\alpha}$ are only approximate, the actual overall probabilities of Types I and II errors being bounded above by $\frac{\alpha}{1 - \beta}$ and $\frac{\beta}{1 - \alpha}$, respectively; these are not generally the least upper bounds. What is desired is a sequential procedure (or sequential procedures) with exactly attainable frequency characteristics when the null hypothesis is true and capable of attaining exactly any given power against any given alternative hypothesis hopefully by setting an upper bound on the sample size. Burman (13), Epstein and Sobel (20), Barraclough and Page (9) and English statisticians (Anscombe, Armitage, Barnard, et. al.) made contributions towards determining exact frequency characteristics and sampling plans for Wald's original SPRT, and Epstein (19), Woodall and Kurkjian (53), Burnett (14) and Aroian (4, 5) were among those investigating exact characteristics of truncations of Wald's SPRT in life testing with an exponential distribution, these latter efforts being amenable to generalizations to other distributions and arbitrary test boundaries.

Armitage, McPherson and Rowe (3) and McPherson and Armitage (41) have investigated exact frequency characteristics of a simple and natural method they propose for sequential hypothesis testing on accumulating data, firstly when the null hypothesis is true and again when it is not true. Their publications contain numerical results for the cases of the underlying distribution of the test statistic being binomial, normal and exponential each against a two-sided alternative. The results were used to formulate proposals for sequential sampling plans in the two-tailed binomial and normal cases. Their methods will here be examined with the following purposes in mind:

(i) Extending their results - to one-tailed cases in particular

(ii) Examining a sequential estimation procedure and associated inferential problems.

(iii) The inferential base of the methods employed will be criticised and alternative modes of inference proposed and criticised.

CHAPTER II

FREQUENCY CHARACTERISTICS OF A METHOD OF SEQUENTIAL HYPOTHESIS AND SIGNIFICANCE TESTING WHEN THE NULL HYPOTHESIS IS TRUE

As Armitage, McPherson and Rowe (3, page 235) have stated,

> The general effect of performing repeated significance tests at different stages during the accumulation of a body of data is well known. If the null hypothesis is true and if each significance test is performed at the same nominal level, the probability that at some stage or another the test criterion is significant may be substantially greater than the nominal value.

They consider problems associated with testing for the significance of accumulating observations using fixed-sample-size procedures. Questions arising naturally are:

(a) What is the probability of obtaining a result "significant" at a certain nominal level within the first (say) 50 tests?

(b) Does the probability of obtaining a "significant" result reach a "noticeably high" level only after a "very large" number of tests?

(c) What is the effect of repeated tests when the null hypothesis is not true?

The purpose of the paper (3) and of McPherson and Armitage's later publication (41) was to answer some of these questions. Sequential observations from three distributional forms were considered: binomial, normal and exponential. The results were used to formulate proposals for sequential sampling plans

> which can be interpreted either from the frequency point of view, with specified probabilities of errors, or as repeated significance tests at a specified level, or perhaps as having a stopping rule defined... (3, page 236).

Two-tailed Normal Case

Armitage, McPherson and Rowe (3) considered the following:

An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are (under the null hypothesis) independently and normally distributed with zero mean and unit variance. After each observation the experimenter uses the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i \tag{2.1}$$

to decide whether to continue sampling. Sampling stops (with the rejection of the null hypothesis) the first time

$$|s_n| > z_\alpha \sqrt{n} \tag{2.2}$$

where for some $\alpha \,\varepsilon (0, \frac{1}{2})$,

$$P(|Z| > z_\alpha) = 2\alpha, \quad Z \sim N(0, 1).$$

The value of n at which the experiment stops will be denoted by m. The immediate problem is to determine the (cumulative) distribution of random variable M.

Let

$$g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -z_\alpha \leq x \leq z_\alpha,$$

let

$$f_1(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and define

$$g_n(x) = \int_{-z_\alpha\sqrt{n-1}}^{z_\alpha\sqrt{n-1}} g_{n-1}(u) f_1(x-u).\, du, \tag{2.3}$$

$$-z_\alpha\sqrt{n} \leq x \leq z_\alpha\sqrt{n},$$

$$n = 2, 3, \ldots .$$

Let P_n denote $P(M \leq n)$, $n = 1, \ldots,$

so $P_1 = 2\alpha;$

then for $n = 2, \ldots,$

$$P_n = 1 - \int_{-z_\alpha\sqrt{n}}^{z_\alpha\sqrt{n}} g_n(x).dx \tag{2.4}$$

$$= P_{n-1} + 2\int_{-z_\alpha\sqrt{n-1}}^{z_\alpha\sqrt{n-1}} g_{n-1}(u)(1-\Phi(z_\alpha\sqrt{n}-u)).du \tag{2.5}$$

where $\Phi(x) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}.dt.$

To simplify the numerical calculation of the P_n's let

$$h_1(x) \equiv e^{-\frac{1}{2}x^2}$$

and

$$h_n(x) = \int_0^{z_\alpha\sqrt{n-1}} h_{n-1}(u)(h_1(x-u) + h_1(x+u)).du, \tag{2.6}$$

$$0 \leq x \leq z_\alpha\sqrt{n},$$

$$n = 2, \ldots .$$

Note (i) $h_1(x) \equiv \sqrt{2\pi} f_1(x)$,

(ii) $h_n(x) = (2\pi)^{\frac{n}{2}} g_n(x), \quad 0 \le x \le z_\alpha \sqrt{n}, \quad n = 2, \ldots,$

(iii) $h_n(0) = 2\int_0^{z_\alpha \sqrt{n-1}} h_{n-1}(u) h_1(u).du,$

and (iv) h_n is an even function when the domain over which it is defined is extended to the entire real line, the definition of h_n extending naturally, $n = 2, \ldots$.

Then (2.4) may be written

$$P_n = 1 - 2(2\pi)^{-\frac{n}{2}} \int_0^{z_\alpha \sqrt{n}} h_n(x).dx, \quad n = 1, \ldots . \qquad (2.7)$$

(i) - (iv) can be used to simplify the computation of the P_n's from (2.6) and (2.7) for any given α. Tables are given in (3).

Note that the experimenter need not necessarily run a test after each observation. Suppose that the predetermined numbers m_i of random observations are made on the normal population between the (i-1)th and ith tests ($i = 1, \ldots$); then letting x_{ij} denote the jth randomly sampled observation between the successive tests ($j = 1, \ldots, m_i$), the experimenter could use the cumulative sum

$$s'_n \equiv \sum_{i=1}^{n} \frac{1}{\sqrt{m_i}} \sum_{j=1}^{m_i} x_{ij}$$

to decide whether to continue sampling. If sampling stops (with the rejection of the null hypothesis) the first time

$$|s'_n| > z_\alpha \sqrt{n},$$

then under the null hypothesis the distribution theory of M (the random variable corresponding to the value of n at which the experiment stops) is as given above. If $m_i = m \; \forall \; i = 1, \ldots,$ then this modified procedure is greatly clarified and the algebra and numerical calculations greatly simplified. If $m_i \neq m_k$ for some $i, k = 1, \ldots,$ then X_{i1} and X_{k1} (for example) will not have "equal weights" in the sequential procedure in the sense that X_{i1} is "diluted" by the factor $\frac{1}{\sqrt{m_i}}$ while X_{k1} is diluted by the factor $\frac{1}{\sqrt{m_k}}$, and these two factors are unequal. Thus unless all the m_i's are equal this alternative procedure "seems unreasonable." However, if all m_i's are equal then essentially the original analysis is applicable.

One-tailed Normal Case

An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are (under the null hypothesis) independently and normally distributed with zero mean and unit variance, and after each observation the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling. Sampling stops (with the rejection of the null hypothesis) the first time $s_n > z_\alpha \sqrt{n}$, where $P(Z > z_\alpha) = \alpha$, $Z \sim N(0, 1)$. The value of n at which the experiment stops will again be denoted by m, and again the immediate problem is to determine the distribution of random variable M.

Let

$$g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \leq z_\alpha,$$

let

$$f_1(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and define

$$g_n(x) = \int_{-\infty}^{z_\alpha\sqrt{n-1}} g_{n-1}(u)f_1(x-u).du, \quad x \leq z_\alpha\sqrt{n}, \quad n = 2, \ldots .$$

Again let

$$h_1(x) \equiv e^{-\frac{1}{2}x^2}$$

and

$$h_n(x) = \int_{-\infty}^{z_\alpha\sqrt{n-1}} h_{n-1}(u)h_1(x-u).du, \quad x \leq z_\alpha\sqrt{n}, \quad n = 2, \ldots .$$

Let P_n again denote $P(M \leq n)$;

then $$P_n = 1 - (2\pi)^{-\frac{n}{2}}\int_{-\infty}^{z_\alpha\sqrt{n}} h_n(x).dx, \; n = 1, \ldots . \tag{2.8}$$

Here $P_1 = \alpha$.

Note (i) $h_1(x) \equiv \sqrt{2\pi}\, f_1(x)$,

(ii) $h_n(x) \equiv (2\pi)^{\frac{n}{2}} g_n(x), \quad n = 2, \ldots,$

(iii) $$h_n(0) = \int_{-\infty}^{z_\alpha\sqrt{n-1}} h_{n-1}(u)h_1(-u).du, \quad n = 2, \ldots$$

$$= \int_{-z_\alpha\sqrt{n-1}}^{\infty} h_{n-1}(-v)h_1(v).dv, \quad v \equiv -u$$

$$= \int_0^{z_\alpha\sqrt{n-1}} h_{n-1}(u)e^{-\frac{1}{2}u^2}.du$$

$$+\int_0^\infty h_{n-1}(-v)e^{-\frac{1}{2}v^2}.dv, \quad n = 2, \ldots,$$

(iv) h_n here is not an even function when the domain over which it is defined is extended to the entire real line, the definition of h_n extending naturally,

$$\text{(v)} \quad h_n(x) = \int_{-z_\alpha\sqrt{n-1}}^{\infty} h_{n-1}(-v)h_1(x+v).dv, \quad v \equiv -u, \quad x \leq z_\alpha\sqrt{n}, \quad n = 2, \ldots$$

$$= \int_0^{z_\alpha\sqrt{n-1}} h_{n-1}(u)e^{-\frac{1}{2}(x-u)^2}.du$$

$$+\int_0^\infty h_{n-1}(-v)e^{-\frac{1}{2}(x+v)^2}.dv, \quad 0 \leq x \leq z_\alpha\sqrt{n}, \quad n = 2, \ldots,$$

and

$$\text{(vi)} \quad h_n(-x) = \int_0^{z_\alpha\sqrt{n-1}} h_{n-1}(u)e^{-\frac{1}{2}(x+u)^2}.du$$

$$+\int_0^\infty h_{n-1}(-v)e^{-\frac{1}{2}(x-v)^2}.dv, \quad v \equiv -u, \quad x > 0, \quad n = 2, \ldots .$$

Then (2.8) may be written

$$P_n = 1 - (2\pi)^{-\frac{n}{2}}\left\{\int_0^{z_\alpha\sqrt{n}} h_n(x).dx + \int_0^\infty h_n(-x).dx\right\}, \quad n = 2, \ldots . \tag{2.9}$$

(i) - (vi) may be used to simplify the computation of the P_n's from (2.9) for any given α. Results are given in Table I.

The basic method was to evaluate the right-hand sides of (iii), (v), (vi) and (2.9) at points on a grid of mesh δ. This was done

TABLE I

P_n's FOR THE ONE-TAILED NORMAL CASE FOR TWO VALUES OF α

n	α = 0.05	α = 0.01
1	0.05000	0.01000
2	0.08008	0.01727
3	0.10105	0.02280
4	0.11706	0.02727
5	0.12997	0.03100
6	0.14076	0.03422
7	0.15001	0.0370
8	0.15811	0.0396
9	0.1653	0.0418
10	0.1718	0.0439
12	0.1830	0.0475
14	0.1925	0.0507
16	0.2008	0.0535
18	0.2080	0.0560
20	0.2145	0.0582
25	0.2282	0.0630

using Simpson's rules (piecewise quadratic or cubic - depending on whether there are three points or four left on the grid). $\delta = 0.1$ was found satisfactory. Special allowance has to be made near the limits of integration where there are incomplete grid-meshes.

Two-tailed Exponential Case

Armitage, McPherson and Rowe (3) considered the following: An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are (under the null hypothesis) independently and exponentially distributed with unit parameter, and after each observation the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling. $2S_n \sim \chi^2(2n)$, $n = 1, \ldots,$ and sampling stops (with the rejection of the null hypothesis) the first time

$$s_n \notin \left[\tfrac{1}{2}\chi^2_{1-\alpha}(2n),\ \tfrac{1}{2}\chi^2_{\alpha}(2n)\right], \text{ where } 0 < \alpha < \tfrac{1}{2}$$

$$\text{and} \int_{\chi^2_\beta(2n)}^{\infty} f_{2n}(x).dx \equiv \beta,$$

f_{2n} being the density (with respect to Lebesgue measure) of a chi-squared random variable with $2n$ degrees of freedom.

Let y_{1n} denote $\frac{1}{2}\chi^2_{1-\alpha}(2n)$, and y_{2n} denote $\frac{1}{2}\chi^2_{\alpha}(2n)$, $n = 1, \ldots$. Again the value of n at which the experiment stops is denoted by m, and again the immediate problem is to find the distribution of random variable M.

Let

$$g_1(x) = e^{-x}, \quad y_{11} \le x \le y_{21},$$

let

$$f_1(x) = \begin{cases} e^{-x}, & x \ge 0, \\ 0 & \text{otherwise}, \end{cases}$$

and for $n = 2, \ldots,$

define

$$g_n(x) = \int_{y_{1,n-1}}^{y_{2,n-1}} g_{n-1}(u) f_1(x-u).du, \quad y_{1n} \le x \le y_{2n}$$

$$= e^{-x} \int_{y_{1,n-1}}^{\min\{x,\, y_{2,n-1}\}} g_{n-1}(u) e^{u}.du, \quad y_{1n} \le x \le y_{2n}.$$

Letting P_n again denote $P(M \le n)$

then $$P_n = 1 - \int_{y_{1n}}^{y_{2n}} g_n(x).dx, \quad n = 1, \ldots .$$

(Obviously $P_1 = 2\alpha$ again.)

<u>Example</u>: $$g_2(x) = e^{-x} \int_{y_{11}}^{\min\{x, y_{21}\}} g_1(u) e^{u}.du$$

$$= \begin{cases} (x - y_{11}) e^{-x}, & y_{12} \le x \le y_{21}, \\ (y_{21} - y_{11}) e^{-x}, & y_{21} \le x \le y_{22} \end{cases} \tag{2.10}$$

and

$$P_2 = 1 - \int_{y_{12}}^{y_{21}} (x - y_{11})e^{-x}.dx - (y_{21} - y_{11})\int_{y_{21}}^{y_{22}} e^{-x}.dx$$

$$= 1 + e^{-y_{21}} + (y_{11} - y_{12} - 1)e^{-y_{12}} + (y_{21} - y_{11})e^{-y_{22}} \tag{2.11}$$

$$= 0.1615836 \quad \text{using } \alpha = 0.05$$

so $y_{11} = \ln\frac{20}{19}$,

$y_{21} = \ln 20$,

$y_{12} = \frac{1}{2} \cdot 0.710723$

and $y_{22} = \frac{1}{2} \cdot 9.48773$.

This example suggests it is simpler to define

$$h_1(x) = 1, \quad x \geq 0,$$

and
$$h_n(x) = \int_{y_{1,n-1}}^{\min\{x,\, y_{2,n-1}\}} h_{n-1}(u).du, \tag{2.12}$$
$$y_{1n} \leq x \leq y_{2n}, \quad n = 2, \ldots,$$

so
$$P_n = 1 - \int_{y_{1n}}^{y_{2n}} h_n(x)e^{-x}.dx, \tag{2.13}$$
$$n = 1, \ldots .$$

Note (i) $h_1(x) = e^x f_1(x), \quad x \geq 0,$

(ii) $h_n(x) \equiv e^x g_n(x), \quad n = 2, \ldots,$ (2.14)

and (iii) h_n is constant on $[y_{2,n-1}, y_{2n}]$, $n = 2, \ldots$.

(2.10), (2.12) - (2.14) and (iii) may be used to facilitate the computations of the P_n's for any given α.

The method was to evaluate the right-hand side of (2.12) at points on a grid of mesh δ, i.e. for

$$u = \lambda_{n-1}\delta(\delta)\mu_{n-1}\delta$$

where $$(\lambda_{n-1} - 1)\delta < y_{1,n-1} \leq \lambda_{n-1}\delta$$

and $$\mu_{n-1}\delta \leq y_{2,n-1} < (\mu_{n-1} + 1)\delta,$$

and at $y_{1,n-1}$, $\frac{1}{2}(y_{1,n-1} + \lambda_{n-1}\delta)$, $\frac{1}{2}((\lambda_n - 1)\delta + y_{1n})$,

y_{1n}, $\frac{1}{2}(y_{1n} + \lambda_n\delta)$, $\frac{1}{2}(\mu_{n-1}\delta + y_{2,n-1})$ and $y_{2,n-1}$.

(By (iii), $h_{n-1}(y_{2,n-1}) = h_{n-1}(\frac{1}{2}(\mu_{n-1}\delta + y_{2,n-1}))$, $n = 2, \ldots$.)

This was done by

(i) the trapezoidal rule (piecewise linear) with $\delta = 0.1$,

(ii) Simpson's rules (piecewise quadratic or cubic - depending whether there are three points or four left on the grid) with $\delta = 0.1$ and 0.05.

Special allowance has to be made near the limits of integration, where there are incomplete grid-meshes. P_n was evaluated from (2.13) by using such methods. These methods are against the advice of Armitage, McPherson and Rowe so comparison of the results given by the above methods with those obtained by their methods is of interest.

Values of y_{1n} and y_{2n} were obtained from tables (47) and using the algorithm of Wilson and Hilferty (52) which was given by Thompson (47) and again by Merrington (42), who checked its accuracy. Armitage, McPherson and Rowe expressed a hope of using such an algorithm

TABLE II

P_n's FOR THE TWO-TAILED EXPONENTIAL CASE FOR VARIOUS VALUES OF 2α

n	2α = 0.10	0.05	0.02	0.01
1	0.10000	0.05000	0.02000	0.01000
2	0.16158	0.08381	0.03468	0.01766
3	0.20402	0.10841	0.04596	0.02375
4	0.23599	0.12753	0.05502	0.02874
5	0.26151	0.14313	0.06258	0.03295
6	0.28267	0.15628	0.06905	0.03660
7	0.30071	0.16764	0.07471	0.03981
8	0.31640	0.17763	0.07974	0.04268
9	0.33027	0.18654	0.08426	0.04528
10	0.34268	0.19458	0.08837	0.04765
12	0.36410	0.20862	0.09563	0.05185
14	0.38211	0.22060	0.10188	0.05550
16	0.39761	0.23102	0.10735	0.05869
18	0.41118	0.24025	0.11224	0.06157
20	0.42322	0.24853	0.11667	0.06419
25	0.44837	0.26608	0.12615	0.06982
30	0.46852	0.28042	0.13401	0.07451
35	0.48524	0.29252	0.14071	0.07854
40	0.49947	0.30296	0.14654	0.08207
45	0.51183	0.31214	0.15172	0.08521
50	0.52271	0.32032	0.15636	0.08804
60	0.54116	0.33439	0.16446	0.09299
70	0.55637	0.34619	0.17130	0.09720
80	0.56925	0.35634	0.17725	0.10087
90	0.58038	0.36522	0.18250	0.10412
100	0.59016	0.37310	0.18720	0.10704
120	0.60665	0.38661	0.19532	0.11211
140	0.62018	0.39790	0.20219	0.11641
160	0.63162	0.4076	0.2080	0.12019

TABLE III

INVERSE NOMINAL SIGNIFICANCE LEVELS $2\alpha(n,L_0)$ IN THE TWO-TAILED EXPONENTIAL CASE FOR GIVEN TERMINAL VALUES OF n TO ACHIEVE THE GIVEN OVERALL SIGNIFICANCE LEVEL L_0 (AFTER THE n TESTS)

n	L_0 = 0.1000	0.0500	0.0200	0.0100
2	2α = 0.0602	0.0292	0.0113	0.0056
3	0.0458	0.0219	0.0083	0.0041
4	0.0381	0.0180	0.0067	0.0033
5	0.0333	0.0156	0.0057	0.0028
10	0.0229	0.0108	0.004	
20	0.0166	0.0075		
50	0.0116	0.005		
100	0.009			
150	0.008			

in their future work to reduce the effect of errors due to inaccuracies in the values of y_{1n} and y_{2n}. (For $2\alpha = 0.05$ their program yielded $P_1 = 0.051$!) Results are given in Tables II and III.

Furthering the above example one finds

$$h_3(x) = \begin{cases} \tfrac{1}{2}x^2 - y_{11}x + y_{11}y_{12} - \tfrac{1}{2}y_{12}^2, & y_{13} \le x \le y_{21}, \\ (y_{21} - y_{11})x + y_{11}y_{12} - \tfrac{1}{2}y_{12}^2 - \tfrac{1}{2}y_{21}^2, & y_{21} \le x \le y_{22}, \\ y_{11}y_{12} - y_{11}y_{22} - \tfrac{1}{2}y_{12}^2 - \tfrac{1}{2}y_{21}^2 + y_{21}y_{22}, & y_{22} \le x \le y_{23}, \end{cases} \tag{2.15}$$

and

$$\begin{aligned} P_3 &= 1 + (y_{11} - y_{11}y_{12} + y_{11}y_{13} + \tfrac{1}{2}y_{12}^2 - y_{13} - \tfrac{1}{2}y_{13}^2 - 1)e^{-y_{13}} \\ &\quad + e^{-y_{21}} + (y_{21} - y_{11})e^{-y_{22}} \\ &\quad + (y_{11}y_{12} - y_{11}y_{22} - \tfrac{1}{2}y_{12}^2 - \tfrac{1}{2}y_{21}^2 + y_{21}y_{22})\, e^{-y_{23}} \end{aligned} \tag{2.16}$$

$= 0.2040170$ using $\alpha = 0.05$,

$y_{11}, y_{12}, y_{21}, y_{22}$ as before,

$$y_{13} = \tfrac{1}{2} \cdot 1.635383$$

and $$y_{23} = \tfrac{1}{2} \cdot 12.59159.$$

Bhate (10) derived formulas analogous to (2.10) and (2.15) using an unnecessarily complicated method, namely inversion of characteristic functions. He exemplified this method in the case where the cut-off boundaries in each tail are linear in n (the number of observations in the cumulative sum) and parallel, but states that his method can be used even when these boundaries are not linear and parallel but "the computations involved will be much more complicated." Using the method of the above example all these objections are relatively easily overcome.

Bhate does however raise an interesting application of the surrounding theory to a class of problems which can be greatly broadened as follows. Suppose one is "investigating" (i.e. intending to test a null hypothesis about) the variance of a normal distribution with known mean μ. (Bhate considered only this case, but the case where the mean is unknown will also be mentioned here soon.) Moreover, suppose the null hypothesis is H_0: $\sigma^2 = \sigma_0^2$ and is to be tested against the two-sided alternative H_A: $\sigma^2 \neq \sigma_0^2$ using a sequential procedure. Randomly sample <u>two</u> observations at a time (i.e. between successive sequential tests) from the normal population. For $k = 1, \ldots,$ and $X_1, \ldots, X_{2k}$ independent and identically distributed $N(0, \sigma^2)$,

$$T_k \equiv \frac{1}{\sigma_0^2}((X_{2k-1} - \mu)^2 + (X_{2k} - \mu)^2) \qquad (2.17)$$

$$\sim \chi^2(2) \qquad \text{under } H_0;$$

i.e. random variables T_k are independent and exponentially distributed with parameter $\lambda = \frac{1}{2}$. Thus, making two observations at a time on this normal population is equivalent to making a single observation on this exponential population.

Hence to test H_0 against H_A using this sequential procedure one <u>could</u> preselect a number of observations to make on the exponential population, say 50 (so this then requires that 50 pairs of observations be taken from the underlying normal population), and preselect an overall size for the test, say 0.10, then keep sampling until either

(i) $\sum_{k=1}^{n} T_k < \chi^2_{1-a}(2n)$ or $\sum_{k=1}^{n} T_k > \chi^2_{a}(2n)$

for some $n = 1, \ldots, 50$, where, by interpolation in Table 3 of Armitage, McPherson and Rowe's publication (3), $a = 0.0116$; in this case H_0 is rejected;

or (ii) the fifty pairs of observations have been sampled from the normal population, in which case H_0 is not rejected (but H_A need not necessarily be rejected either).

No claim is made that this procedure is optimal in any sense, just that it is an illuminating and apparently reasonable application of the surrounding theory. (Stein (46) has stated that "It is difficult even to formulate a definition of an optimal among sequential tests of a hypothesis against multiple alternatives.")

Of course in practice one may be tempted to

(i) stop without rejecting H_0 before taking fifty pairs of observations from the normal population if there seems little likelihood of rejecting H_0 before observing the fiftieth pair,

or (ii) continue random sampling beyond the preset limit of fifty pairs of observations from the normal population if rejection of H_0 at the preset overall size of the test seems imminent after the fiftieth pair of normal observations has been sampled.

This latter procedure is of course objectionable from many points of view, among these objections being the fact that this procedure increases the overall size of the test beyond the preset overall size. Again in practice one may prefer to run sequential tests only after every two pairs of observations have been randomly sampled from the normal population, in which case a new problem arises - that of the "two-tailed $\chi^2(4)$ case." Obviously there is no limit to the natural theoretical extensions here. Another approach would be to not reject H_0 until two or three sequential tests had been judged "significant."

In the case where the mean of the population is unknown one could take three random samples from the normal population before applying the first test of the sequence; then for $n = 1, \ldots,$ and

$X_0, X_1, \ldots, X_{2n}$ i.i.d. $N(\mu, \sigma^2)$,

$$T'_n \equiv \frac{1}{\sigma_0^2} \sum_{i=0}^{2n} (X_1 - \overline{X}^{(n)})^2, \quad \overline{X}^{(n)} \equiv \frac{1}{2n+1} \sum_{i=0}^{2n} X_i \qquad (2.18)$$

$$\sim \chi^2(2n) \qquad \text{under } H_0.$$

Note that $T'_{n+1} \geq T'_n$ with equality if and only if

$$X_{2n+1} = \overline{X}^{(n)} = X_{2(n+1)}$$

(so $T'_{n+1} \gneq T'_n$ almost surely), $n = 1, \ldots$.

Hence $T'_{n+1} - T'_n \sim \chi^2(2)$

and $T'_{n+1} - T'_n$ and T'_n are independent, $n = 1, \ldots$.

Thus H_0 can be tested against

H_A using the previous test procedure (when the mean was assumed known) with $\sum_{k=1}^{n} T_k$ replaced by T_n'.

Again one may be interested in testing H_0 against

$$H_1: \sigma^2 > \sigma_0^2 , \quad \text{or}$$

$$H_2: \sigma^2 < \sigma_0^2 ,$$

where the mean may be known or unknown. Extending the above procedures in the obvious manner, to test H_0 against H_1 one could preselect a number of pairs of observations to be randomly sampled from the normal population (the first "pair" being <u>three</u> observations when the mean is unknown) and preselect an overall size for the test procedure, then keep randomly sampling until either

(i) $\sum_{k=1}^{n} T_k > \chi_a^2(2n)$ or $T_n' > \chi_a^2(2n)$ for some integral $n \leq p$, a here being obtained by interpolating in appropriate tables* (different from the Table referenced above) and is such that the overall size of the sequential test procedure is the preselected value; in this case H_0 is rejected and H_1 accepted;

or (ii) the p pairs of observations have been sampled from the population, in which case H_0 is <u>not</u> rejected (but H_1 need not necessarily be rejected either).

Similarly to test H_0 against H_2 one could "legitimately" reject H_0 if and only if

$$\sum_{k=1}^{n} T_k < \chi_{1-a}^2(2n) \quad \text{or} \quad T_k' < \chi_{1-a}^2(2n) \text{ for some integral } n \leq p,$$

a here being different from the a's in the above test criteria but again being derived by the same method (interpolation in the appropriate tables*) and tailored to suit the same purpose (making the overall test procedure the preselected overall size).

*The objection now is that the "appropriate tables" from which a is to be determined do not exist to this point; i.e. the test criteria necessitate new tables. The one-tailed exponential cases (from which these tables will come) will soon be discussed.

Comparison of Two-tailed Normal and Two-tailed Exponential Results

Comparing the two-tailed normal table given by Armitage, McPherson and Rowe (3) with the two-tailed exponential results, two general trends are to be observed for each of the chosen values of 2α:

(i) For "smaller" values of n ($n = 2, \ldots, 20$) the P_n's in the normal table are less than the corresponding P_n's in the exponential case. This means that for a maximum number of these sequential tests in this "lower" range the nominal significance level at which each test is to be conducted to achieve a given overall significance level (after the maximum number of tests) is greater in the normal case than in exponential testing. This in turn suggests that if an experimenter plans to use a sequential testing procedure described above then, assuming the test statistics obtained from the experiment are continuous and amenable to conversion to normal or exponential statistics of equal significance level, it is preferable to convert them to normal test statistics.

(ii) For "larger" values of n (greater than 60) the opposite is true. This may be a manifestation of the asymptotic optimality of Fisher's method (38, 39) in which case some partial answers may be provided as to just how large a sample size of independent test statistics is necessary before using Fisher's method as more powerful than other methods of combination.

One-tailed Exponential Cases

Right tail. An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are (under the null hypothesis) independently and exponentially distributed with unit parameter, and after each observation the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling. As with the two-tailed exponential case, $2S_n \sim \chi^2(2n)$, $n = 1, \ldots$. Sampling stops (with the rejection of the null hypothesis) the first time

$$s_n > \tfrac{1}{2}\chi^2_\alpha(2n), \quad \text{where} \quad 0 < \alpha < 1 \quad \text{and}$$

$$\int_0^{\chi^2_\alpha(2n)} f_{2n}(x).dx = 1 - \alpha, \quad f_{2n} \text{ the density}$$

(with respect to Lebesgue measure) of a chi-squared random variable with $2n$ degrees of freedom.

Let y_n denote $\tfrac{1}{2}\chi^2_\alpha(2n)$, $n = 1, \ldots$. Again the value of n at which the experiment stops will be denoted by m, and again the immediate problem is to determine the distribution of random variable M.

Let
$$g_1(x) = e^{-x}, \quad 0 \leq x \leq y_1,$$

$$f_1(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0 & \text{elsewhere,} \end{cases}$$

and for $n = 2, \ldots,$

define
$$g_n(x) = e^{-x}\int_0^{\min\{x,\, y_{n-1}\}} g_{n-1}(u)e^{u}.du, \quad 0 \leq x \leq y_n,$$

and
$$P_n \equiv P(M \leq n)$$

$$= 1 - \int_0^{y_n} g_n(x).dx.$$

Here
$$P_1 = \alpha.$$

Again define

$$h_1(x) = 1, \quad x \geq 0,$$

and define

$$h_n(x) = \int_0^{\min\{x,\, y_{n-1}\}} h_{n-1}(u).du, \quad 0 \leq x \leq y_n, \quad n = 2, \ldots, \tag{2.19}$$

so
$$P_n = 1 - \int_0^{y_n} h_n(x)e^{-x}.dx, \quad n = 2, \ldots . \tag{2.20}$$

Again note (i) $h_1(x) = e^{x}f_1(x), \quad x \geq 0,$

(ii) $h_n(x) \equiv e^{x}g_n(x), \quad n = 2, \ldots,$ (2.14)

and (iii) h_n is constant on $[y_{n-1}, y_n]$, $n = 2, \ldots .$

Also, for $n = 2, \ldots,$ $P(M=n)$ is the probability that sampling continues through the first $n-1$ samples and stops at the nth sample, so that $P(M=n)$ is the probability that $S_{n-1} \in [0, y_{n-1}]$ (this probability being measured by the integral of g_{n-1} over this interval) <u>and</u> $X_n > y_n - S_{n-1}$. Mathematically,

$$P(M=n) = \int_0^{y_{n-1}} g_{n-1}(u) \int_{y_n - u}^{\infty} f_1(x).dx.du,$$

so by (2.14),

$$P_n = P_{n-1} + e^{-y_n} \int_0^{y_{n-1}} h_{n-1}(u).du \ . \qquad (2.21)$$

(2.19) and (iii) may be used to facilitate the computations of the P_n's from (2.20) or (2.21) for any given α.

The final program used in calculating the P_n's in this case utilizes (2.20) with grid-mesh $\delta = 0.05$ to $n = 25$ and (2.21) with $\delta = 0.1$ thereafter. Results are given in Tables IV and V.

Again, if the domain over with g_n and h_n are defined is extended to $[0, \infty)$, the definitions of g_n and h_n extending naturally, then

$$P(M=n) = \int_{y_n}^{\infty} g_n(x).dx,$$

so

$$P_n = P_{n-1} + \int_{y_n}^{\infty} h_n(x)e^{-x}.dx. \qquad (2.22)$$

TABLE IV

P_n's IN THE RIGHT-TAILED EXPONENTIAL CASE FOR VARIOUS VALUES OF α

n	α = 0.05	0.025	0.01	0.005
1	0.05000	0.02500	0.01000	0.00500
2	0.07608	0.03904	0.01603	0.00814
3	0.09401	0.04905	0.02049	0.01052
4	0.10774	0.05691	0.02408	0.01246
5	0.11889	0.06340	0.02709	0.01410
6	0.12827	0.06894	0.02970	0.01554
7	0.13638	0.07379	0.03201	0.01682
8	0.14352	0.07809	0.03408	0.01798
9	0.14989	0.08196	0.03596	0.01903
10	0.15566	0.08549	0.03768	0.02000
12	0.16574	0.09172	0.04075	0.02173
14	0.17435	0.09709	0.04342	0.02325
16	0.18188	0.10182	0.04579	0.02460
18	0.18855	0.10605	0.04792	0.02582
20	0.19454	0.10987	0.04987	0.02694
25	0.20729	0.11808	0.05408	0.02938
30	0.21774	0.12488	0.05761	0.03144
35	0.22658	0.13069	0.06066	0.03323
40	0.23424	0.13577	0.06333	0.03480
45	0.24100	0.14027	0.06572	0.03622
50	0.24703	0.14431	0.06788	0.03750
60	0.25745	0.15134	0.07166	0.03975
70	0.26622	0.15731	0.07490	0.04169
80	0.27380	0.16250	0.07773	0.04338
90	0.28046	0.16708	0.08024	0.04490
100	0.28639	0.17119	0.08250	0.04626
120	0.29659	0.17830	0.08644	0.04865
140	0.30516	0.18432	0.08980	0.05069
160	0.31256	0.18953	0.0927	0.0525

TABLE V

INVERSE NOMINAL SIGNIFICANCE LEVELS $\alpha(n,L_0)$ IN THE RIGHT-TAILED EXPONENTIAL CASE FOR GIVEN TERMINAL VALUES OF n TO ACHIEVE THE GIVEN OVERALL SIGNIFICANCE LEVEL L_0 (AFTER THE n TESTS)

n	L_0 = 0.0500	0.0250	0.0100	0.0050
2	α = 0.0323	0.0158	0.0062	0.0030
3	0.0255	0.0123	0.0048	0.0023
4	0.0218	0.0104	0.0040	0.0019
5	0.0193	0.0092	0.0035	0.0017
10	0.0136	0.0064	0.002	
20	0.0100	0.0046		
50	0.0069	0.003		
100	0.0055			
150	0.0048			

Examples: $h_2(x) = \min\{x, y_1\}, \quad 0 \leq x \leq y_2,$

so by (2.20),

$$P_2 = 1 - \int_0^{y_1} xe^{-x}.dx - y_1 \int_{y_1}^{y_2} e^{-x}.dx$$

$$= e^{-y_1} + y_1 e^{-y_2}$$

$$= P_1 + y_1 e^{-y_2}, \tag{2.23}$$

which is what (2.22) gives directly and also (2.21)

$= 0.07607766$ using $= 0.05$

so $y_1 = \ln 20$

and $y_2 = \frac{1}{2} \cdot 9.48773.$

$$h_3(x) = \begin{cases} \frac{1}{2}x^2, & 0 \leq x \leq y_1, \\ y_1 x - \frac{1}{2}y_1^2, & y_1 \leq x \leq y_2, \\ y_1(y_2 - \frac{1}{2}y_1), & y_2 \leq x \leq y_3, \end{cases}$$

and by (2.20),

$$P_3 = e^{-y_1} + y_1 e^{-y_2} + y_1(y_2 - \tfrac{1}{2}y_1)e^{-y_3}$$

$$= P_2 + y_1(y_2 - \tfrac{1}{2}y_1)e^{-y_3}, \tag{2.24}$$

which is what (2.22) gives directly and also (2.21)

$= 0.0940094$ using $\alpha = 0.05$, y_1, y_2 as above

and $y_3 = \frac{1}{2} \cdot 12.59159.$

By observing the pattern developing in the above calculations P_4 may be postulated to be

$$P_3 + y_1(\tfrac{1}{6}\, y_1^{\,2} - \tfrac{1}{2}y_1y_3 - \tfrac{1}{2}y_2^2 + y_2y_3)e^{-y_4} \tag{2.25}$$

$= 0.1077401$ using $\alpha = 0.05$, y_1, y_2, y_3 as before and $y_4 = \frac{1}{2} \cdot 15.50732$,

and, further, P_5 may be postulated to be

$$P_4 + y_1\left[(\tfrac{1}{6}\, y_1^2 - \tfrac{1}{2}y_1y_3 - \tfrac{1}{2}y_2^2 + y_2y_3)y_4 + \tfrac{1}{4}y_1y_3^2 + \tfrac{1}{6}\, y_2^3 - \tfrac{1}{2}y_2^3 - \tfrac{1}{24}\, y_1^3\right]e^{-y_5} \tag{2.26}$$

$= 0.1188853$ using $\alpha = 0.05$, $y_1, \ldots, y_4$ as before and $y_5 = \frac{1}{2} \cdot 18.30705$.

<u>Left tail</u>. An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are (under the null hypothesis) independently and exponentially distributed with unit parameter, and after each observation the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling. As with the previous exponential cases, $2S_n \sim \chi^2(2n)$, $n = 1, \ldots$. Sampling stops (with the rejection of the null hypothesis) the first time

$$s_n < \tfrac{1}{2}\, \chi^2_{1-\alpha}(2n), \quad \text{where} \quad 0 < \alpha < 1 \quad \text{and}$$

$$\int_{\chi^2_{1-\alpha}(2n)}^{\infty} f_{2n}(x).dx = 1 - \alpha,$$

f_{2n} the density (with respect to Lebesgue measure) of a chi-squared random variable with $2n$ degrees of freedom.

Let y_n here denote $\frac{1}{2}\chi^2_{1-\alpha}(2n)$, $n = 1, \ldots$. Again the value of n at which the experiment stops will be denoted by m, and again the immediate problem is to determine the distribution of random variable M.

Let
$$g_1(x) = e^{-x}, \quad x \geq y_1,$$

let
$$f_1(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0 & \text{elsewhere,} \end{cases}$$

and for $n = 2, \ldots,$

define
$$g_n(x) = e^{-x}\int_{y_{n-1}}^{x} g_{n-1}(u)e^{u}.du, \qquad x \geq y_n,$$

and
$$P_n \equiv P(M \leq n)$$

$$= 1 - \int_{y_n}^{\infty} g_n(x).dx,$$

$$P_1 = \alpha.$$

Again let
$$h_1(x) = 1, \quad x \geq 0,$$

and define
$$h_n(x) = \int_{y_{n-1}}^{x} h_{n-1}(u).du, \quad x \geq y_n \quad n = 2, \ldots, \tag{2.27}$$

so $$P_n = 1 - \int_{y_n}^{\infty} h_n(x)e^{-x}.dx, \qquad n = 2, \ldots . \tag{2.28}$$

Again note (i) $h_1(x) = e^x f_1(x), \quad x \geq 0,$

and (ii) $h_n(x) \equiv e^x g_n(x), \quad n = 2, \ldots .$ (2.14)

Also, using an argument equivalent to that given in the right-tailed exponential case, for $n = 2, \ldots,$ $P(M=n)$ is the probability that sampling continues through the first $n-1$ observations and stops at the nth observation, so that $P(M=n)$ is the probability that $S_{n-1} \geq y_{n-1}$ (this probability being measured by the integral of g_{n-1}) <u>and</u> $X_n < y_n - S_{n-1}$,

i.e. $$P(M=n) = \int_{y_{n-1}}^{y_n} g_{n-1}(u) \int_0^{y_n - u} f_1(x).dx.du,$$

so by (2.14),

$$P_n = P_{n-1} + \int_{y_{n-1}}^{y_n} h_{n-1}(u)(e^{-u} - e^{-y_n}).du . \tag{2.29}$$

Again, as in the right-tailed exponential case, if the domain over which g_n and h_n are defined is extended to $[y_{n-1}, \infty)$, the definitions of these functions extending naturally, then

$$P(M=n) = \int_{y_{n-1}}^{y_n} g_n(x).dx,$$

so $$P_n = P_{n-1} + \int_{y_{n-1}}^{y_n} h_n(x)e^{-x}.dx . \tag{2.30}$$

(2.27) (with domain of definition of h_n extended to $[y_{n-1},\infty)$) is used to facilitate the computations of the P_n's from (2.29) or (2.30) for any given α.

The final program used in calculating the P_n's in this case utilizes only (2.29) with grid-mesh $\delta = 0.05$ to $n = 25$ and $\delta = 0.1$ thereafter. Results are given in Tables VI and VII.

<u>Examples</u>:

Here $h_2(x) = x - y_1, \quad x \geq y_1,$

so by (2.30),

$$P_2 = P_1 + \int_{y_1}^{y_2} xe^{-x}.dx - y_1(e^{-y_1} - e^{-y_2})$$

$$= P_1 + (y_1 - y_2 - 1)e^{-y_2} + e^{-y_1}, \quad \text{which is what (2.29) gives directly}$$

$$\equiv 1 + (y_1 - y_2 - 1)e^{-y_2} \qquad (2.31)$$

(since $P_1 + e^{-y_1} \equiv 1$),

which is what (2.28) gives

$= 0.08595249$ using $\alpha = 0.05$

so $y_1 = \ln\frac{20}{19}$

and $y_2 = \frac{1}{2} \cdot 0.710723$.

$$h_3(x) = \int_{y_2}^{x} (u - y_1).du, \quad x \geq y_2$$

$$= \tfrac{1}{2}x^2 - y_1x + y_2(y_1-\tfrac{1}{2}y_2), \quad x \geq y_2,$$

TABLE VI

P_n's FOR THE LEFT-TAILED EXPONENTIAL CASE
FOR VARIOUS VALUES OF α

n	α = 0.05	0.025	0.01	0.005
1	0.05000	0.02500	0.01000	0.00500
2	0.08595	0.04487	0.01866	0.00952
3	0.11099	0.5957	0.02550	0.01323
4	0.12977	0.07097	0.03100	0.01629
5	0.14468	0.08020	0.03555	0.01886
6	0.15699	0.08794	0.03943	0.02107
7	0.16744	0.09458	0.04280	0.0230
8	0.17649	0.10039	0.04578	0.02474
9	0.18446	0.10554	0.04845	0.02629
10	0.19158	0.11018	0.05086	0.02769
12	0.20385	0.11822	0.05508	0.03017
14	0.21415	0.12504	0.05869	0.03231
16	0.22301	0.13094	0.06182	0.03416
18	0.23077	0.13614	0.06461	0.03581
20	0.23767	0.14081	0.06713	0.03732
25	0.25213	0.15064	0.07248	0.04053
30	0.26378	0.15865	0.07687	0.04319
35	0.27350	0.16538	0.08060	0.04545
40	0.28184	0.17119	0.08384	0.04742
45	0.28911	0.17629	0.08669	0.04917
50	0.29556	0.18083	0.08924	0.05073
60	0.30660	0.18865	0.09368	0.05346
70	0.31579	0.19521	0.09741	0.05577
80	0.32366	0.20086	0.10064	0.05777
90	0.33052	0.20581	0.10349	0.05954
100	0.33660	0.21021	0.10603	0.06112
120	0.34699	0.21777	0.11042	0.06385
140	0.35565	0.22411	0.11411	0.06616
160	0.36315	0.22956	0.11731	0.06816
180		0.23435	0.12012	0.06993
200			0.12262	0.07151

TABLE VII

INVERSE NOMINAL SIGNIFICANCE LEVELS $\alpha(n,L_0)$ IN THE LEFT-TAILED EXPONENTIAL CASE FOR GIVEN TERMINAL VALUES OF n TO ACHIEVE GIVEN OVERALL SIGNIFICANCE LEVEL L_0 (AFTER THE n TESTS)

n	L_0 = 0.0500	0.0250	0.0100	0.0050
2	α = 0.0280	0.0134	0.0053	0.0026
3	0.0206	0.0098	0.0037	0.0018
4	0.0169	0.0079	0.0030	0.0014
5	0.0146	0.0068	0.0025	0.0012
10	0.0098	0.0044		
20	0.0070	0.003		
50	0.0049			
100	0.004			
150	0.003			

so by (2.30),

$$P_3 = P_2 - (\tfrac{1}{2}y_3^2 + y_3 + 1 - y_1(y_3 + 1) + y_2(y_1 - \tfrac{1}{2}y_2))e^{-y_3}$$

$$+ (y_2 - y_1 + 1)e^{-y_2}$$, which is what (2.29) yields more readily

$$\equiv 1 - (\tfrac{1}{2}y_3^2 + y_3 + 1 - y_1(y_3+1) + y_2(y_1-\tfrac{1}{2}y_2)).e^{-y_3} \quad (2.32)$$

by (2.31), and is what (2.28) gives

$= 0.1109857$ using $\alpha = 0.05$, y_1, y_2 as before and $y_3 = \tfrac{1}{2} \cdot 1.635383$.

By (2.29),

$$P_4 = P_3 + \int_{y_3}^{y_4} (\tfrac{1}{2}u^2 - y_1u + y_2(y_1 - \tfrac{1}{2}y_2)) \; (e^{-u} - e^{-y_4}).du$$

$$= P_3 - (\tfrac{1}{2}y_4^2 + y_4 + 1 - y_1(y_4 + 1) + y_2(y_1 - \tfrac{1}{2}y_2)(1 + y_4 - y_3) + \tfrac{1}{6}(y_4^3 - y_3^3) - \tfrac{1}{2}y_1(y_4^2 - y_3^2))e^{-y_4}$$

$$+ (\tfrac{1}{2}y_3^2 + y_3 + 1 - y_1(y_3 + 1) + y_2(y_1 - \tfrac{1}{2}y_2))e^{-y_3}$$

$$\equiv 1 - (\tfrac{1}{2}y_4^2 + y_4 + 1 - y_1(y_4 + 1) + y_2(y_1 - \tfrac{1}{2}y_2)(1 + y_4 - y_3) + \tfrac{1}{6}(y_4^3 - y_3^3) - \tfrac{1}{2}y_1(y_4^2 - y_3^2))e^{-y_4} \quad (2.33)$$

by (2.32)

$= 0.1297729$ using $\alpha = 0.05$, y_1, y_2, y_3 as before and $y_4 = \tfrac{1}{2} \cdot 2.732637$.

$$h_4(x) = \int_{y_3}^{x} (\tfrac{1}{2}u^2 - y_1u + y_2(y_1 - \tfrac{1}{2}y_2)).du, \quad x \geq y_3$$

$$= \tfrac{1}{6}x^3 - \tfrac{1}{2}y_1x^2 + y_2(y_1 - \tfrac{1}{2}y_2)x$$

$$- (\tfrac{1}{6}y_3^3 - \tfrac{1}{2}y_1y_3^2 + y_2(y_1 - \tfrac{1}{2}y_2)y_3), \quad x \geq y_3,$$

so by (2.29) and observing the pattern developing in the above calculations, P_5 may be postulated to be

$$1 - (\tfrac{1}{6}y_5^3 + \tfrac{1}{2}y_5^2 + y_5 + 1 - y_1(\tfrac{1}{2}y_5^2 + y_5 + 1) + y_2(y_1 - \tfrac{1}{2}y_2)(y_5+1)$$

$$- (\tfrac{1}{6}y_3^3 - \tfrac{1}{2}y_1y_3^2 + y_2(y_1 - \tfrac{1}{2}y_2)y_3)(1 + y_5 - y_4) + \tfrac{1}{24}(y_5^4 - y_4^4) \qquad (2.34)$$

$$- \tfrac{1}{6}y_1(y_5^3 - y_4^3) + \tfrac{1}{2}y_2(y_1 - \tfrac{1}{2}y_2)(y_5^2 - y_4^2))e^{-y_5}$$

$= 0.1446847$ using $\alpha = 0.05$, $y_1, \ldots, y_4$ as before and $y_5 = \frac{1}{2} \cdot 3.940297$.

CHAPTER III

POWER OF THE METHOD OF SEQUENTIAL TESTING

Power of the Method in the Two-tailed Normal Case

McPherson and Armitage (41) considered the following:

An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are independently and normally distributed with mean μ and unit variance. After each observation the experimenter uses the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i \tag{3.1}$$

to decide whether to continue sampling. Sampling stops (with the rejection of the null hypothesis H_0: $\mu = 0$) the first time

$$|s_n| > z_\alpha \sqrt{n}\,, \quad \text{where} \quad \Phi(z_\alpha) = 1 - \alpha \quad \forall\, \alpha \in (0, \tfrac{1}{2}). \tag{3.2}$$

Again the value of n at which the experiment stops will be denoted by m and again the immediate problem is to determine the distribution of random variable M.

Letting

$$g_1(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2},$$

define

$$g_n(x) \equiv \int_{-z_\alpha\sqrt{n-1}}^{z_\alpha\sqrt{n-1}} g_{n-1}(u) g_1(x-u).du, \tag{3.3}$$

$$n = 2, \ldots .$$

Letting P_n again denote $P(M \leq n)$, then for $n = 1, \ldots,$

$$P_n = 1 - \int_{-z_\alpha\sqrt{n}}^{z_\alpha\sqrt{n}} g_n(x).dx \,. \tag{3.4}$$

The probability of being absorbed in the upper boundary <u>at</u> the nth observation is given by

$$Q_n = \int_{z_\alpha\sqrt{n}}^{\infty} g_n(x).dx \tag{3.5}$$

and similarly for the lower boundary

$$R_n = \int_{-\infty}^{-z_\alpha\sqrt{n}} g_n(x).dx, \qquad n = 1, \ldots \,. \tag{3.6}$$

Note (i): $\sum_{i=1}^{n}(Q_i + R_i) = P_n, \quad n = 1, \ldots$ (3.7)

(not $1 - P_n$ as given by McPherson and Armitage (41) in their Appendix).

(3.7) can be used to check the accuracy and precision of the numerical computations of the P_n's, Q_n's and R_n's. To simplify and facilitate these computations let

$$h_1(x) \equiv e^{-\frac{1}{2}(x-\mu)^2} \tag{3.8}$$

and

$$h_n(x) \equiv \int_{-z_\alpha\sqrt{n-1}}^{z_\alpha\sqrt{n-1}} h_{n-1}(u)h_1(x-u).du, \quad n = 2, \ldots \,. \tag{3.9}$$

Note (ii): $h_n(x) \equiv (2\pi)^{\frac{n}{2}} g_n(x), \quad n = 1, \ldots \,.$

Then (3.4) - (3.6) may be written

$$P_n = 1 - (2\pi)^{-\frac{n}{2}} \int_{-z_\alpha\sqrt{n}}^{z_\alpha\sqrt{n}} h_n(x).dx, \tag{3.10}$$

$$Q_n = (2\pi)^{-\frac{n}{2}} \int_{z_\alpha\sqrt{n}}^{\infty} h_n(x).dx \tag{3.11}$$

and $$R_n = (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{-z_\alpha\sqrt{n}} h_n(x).dx, \quad n = 1, \ldots . \tag{3.12}$$

(3.8) and (3.9) can be used to simplify the computations of the P_n's, Q_n's and R_n's from (3.10) - (3.12) for any given α. Tables are in (41).

Power of the Method in the One-tailed Normal Case

$X_1, \ldots, X_n$ are i.i.d. $N(\mu, 1)$. After each observation the experimenter uses

$$s_n \equiv \sum_{i=1}^{n} x_i$$

to decide whether to continue sampling: sampling stops (with the rejection of H_0: $\mu = 0$) the first time

$$s_n > z_\alpha\sqrt{n}, \quad \Phi(z_\alpha) \equiv 1 - \alpha.$$

Let $$g_1(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \mu)^2}$$

and $$g_n(x) \equiv \int_{-\infty}^{z_\alpha\sqrt{n-1}} g_{n-1}(u)g_1(x-u).du, \quad n = 2, \ldots .$$

Defining

$$h_1(x) \equiv e^{-\frac{1}{2}(x-\mu)^2}$$

and
$$h_n(x) \equiv \int_{-\infty}^{z_\alpha\sqrt{n-1}} h_{n-1}(u)h_1(x-u).du, \qquad n = 2, \ldots,$$

then
$$P_n = P(M \leq n), \quad M \text{ as previously}$$

$$= 1 - (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{z_\alpha\sqrt{n}} h_n(x).dx, \qquad n = 1, \ldots,$$

and
$$Q_n = (2\pi)^{-\frac{n}{2}} \int_{z_\alpha\sqrt{n}}^{\infty} h_n(x).dx$$

is the probability of absorption in the boundary at the nth observation, $n = 1, \ldots$.

Note that $\sum_{i=1}^{n} Q_i = P_n, \quad n = 1, \ldots$. (3.13)

(3.13) can be used to check the accuracy and precision of the numerical computations of the P_n's and Q_n's.

Power of the Method in the Two-tailed Exponential Case

Consider the following:

An experiment consists of a series of observations $x_1, \ldots, x_n$ on random variables which are independently and exponentially distributed with parameter $\lambda \varepsilon (0,\infty)$, i.e.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

After each observation the cumulative sum

$$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling.

$2\,S_n \sim \chi^2(2n)$, $n = 1, \ldots,$ and sampling stops (with the rejection of the null hypothesis H_0: $\lambda = \lambda_0$) the first time

$$s_n \notin \left[\frac{1}{2\lambda_0}\,\chi^2_{1-\alpha}(2n),\ \frac{1}{2\lambda_0}\,\chi^2_{\alpha}(2n)\right], \quad \text{where } 0 < \alpha < \tfrac{1}{2}.$$

Again the value of n at which the experiment stops will be denoted by m and again the immediate problem is to determine the distribution of random variable M.

One is interested in testing H_0: $\lambda = \lambda_0$ against H_A: $\lambda \neq \lambda_0$ where, without loss of generality, λ_0 may be taken as unity (i.e. H_0: $\lambda_0 = 1$: otherwise take $\frac{\lambda}{\lambda_0}$ in place of λ, $\lambda_0 X$ in place of X, and $\lambda_0 x$ in place of x), so that under H_0, $2S_n \sim \chi^2(2n)$, $n = 1, \ldots$. As in Chapter II let y_{1n} denote $\frac{1}{2}\chi^2_{1-\alpha}(2n)$ and let y_2 denote $\frac{1}{2}\chi^2_{\alpha}(2n)$, $n = 1, \ldots$.

Letting $g_1(x) = \lambda e^{-\lambda x}$, $x \geq 0$,

then for $n = 2, \ldots,$ define

$$g_n(x) = \int_{y_{1,n-1}}^{\min\{x, y_{2,n-1}\}} g_{n-1}(u) g_1(x-u).du$$

$$= \lambda e^{-\lambda x} \int_{y_{1,n-1}}^{\min\{x, y_{2,n-1}\}} g_{n-1}(u) e^{\lambda u}.du, \quad x \geq y_{1,n-1}.$$

Letting P_n again denote $P(M \leq n)$ then

$$P_n = 1 - \int_{y_{1n}}^{y_{2n}} g_n(x).dx, \quad n = 1, \ldots .$$

The probability of being absorbed in the upper boundary at the nth observation is

$$Q_n = \int_{y_{2n}}^{\infty} g_n(x).dx$$

and similarly for the lower boundary

$$R_n = \int_{y_{1,n-1}}^{y_{1n}} g_n(x).dx, \quad \text{where} \quad y_{10} \equiv 0.$$

Analogous to Chapter II define

$$h_1(x) = 1, \quad x \geq 0,$$

and $$h_n(x) = \int_{y_{1,n-1}}^{\min\{x, y_{2,n-1}\}} h_n(u).du, \quad x \geq y_{1,n-1}, \quad n = 2, \ldots; \tag{3.14}$$

then $$P_n = 1 - \lambda^n \int_{y_{1n}}^{y_{2n}} h_n(x) e^{-\lambda x}.dx, \tag{3.15}$$

$$Q_n = \lambda^n \int_{y_{2n}}^{\infty} h_n(x) e^{-\lambda x}.dx$$

and $$R_n = \lambda^n \int_{y_{1,n-1}}^{y_{1n}} h_n(x) e^{-\lambda x}.dx, \quad n = 1, \ldots; \; y_{10} \equiv 0. \tag{3.16}$$

Results are given in Tables VIII and IX.

Note (i) $\sum_{i=1}^{n}(Q_i + R_i) = P_n, \quad n = 1, \ldots,$

(ii) $h_n(x) = \lambda^{-n} e^{\lambda x} g_n(x), \quad x \geq y_{1,n-1}, \quad n = 1, \ldots,$

where $y_{10} \equiv 0$,

(iii) h_n is constant on $[y_{2,n-1}, y_{2n}]$, $n = 2, \ldots$.

<u>Examples</u>:

$$P_1 = 1 + \alpha^\lambda - (1 - \alpha)^\lambda$$

and $R_1 = 1 - (1 - \alpha)^\lambda$.

$$h_2(x) = \begin{cases} x - y_{11}, & y_{11} \leq x \leq y_{21}, \\ y_{21} - y_{11}, & x \geq y_{21}, \end{cases} \tag{3.17}$$

$$P_2 = 1 + e^{-\lambda y_{21}} + (\lambda y_{11} - \lambda y_{12} - 1)e^{-\lambda y_{12}} + \lambda(y_{21} - y_{11})e^{-\lambda y_{22}} \tag{3.18}$$

and $$R_2 = e^{-\lambda y_{11}} + (\lambda y_{11} - \lambda y_{12} - 1)e^{-\lambda y_{12}} . \tag{3.19}$$

Similarly,

$$\begin{aligned} P_3 = 1 &+ (\lambda y_{11} - \lambda^2 y_{11} y_{12} + \lambda^2 y_{11} y_{13} + \tfrac{1}{2}\lambda^2 y_{12}^2 - \lambda y_{13} \\ &- \tfrac{1}{2}\lambda^2 y_{13}^2 - 1)e^{-\lambda y_{13}} + e^{-\lambda y_{21}} \\ &+ \lambda(y_{21} - y_{11})e^{-\lambda y_{22}} + \lambda^2(y_{11} y_{12} - y_{11} y_{22} \\ &- \tfrac{1}{2}y_{12}^2 - \tfrac{1}{2}y_{21}^2 + y_{21} y_{22})e^{-\lambda y_{23}} \end{aligned} \tag{3.20}$$

and

TABLE VIII

P_n's FOR THE TWO-TAILED EXPONENTIAL CASE WITH λ=2 AND FOR VARIOUS VALUES OF 2α

n	α = 0.01	0.02	0.05	0.10
1	0.01000	0.02000	0.05000	0.10000
2	0.02686	0.05135	0.11744	0.21289
3	0.04967	0.09015	0.18927	0.31660
4	0.07740	0.13386	0.26095	0.40906
5	0.10925	0.18087	0.33055	0.49082
6	0.14451	0.22998	0.39697	0.56275
7	0.18252	0.28016	0.45954	0.62573
8	0.22263	0.33057	0.51787	0.68066
9	0.26420	0.38048	0.57175	0.72835
10	0.30666	0.42931	0.62115	0.76957
12	0.39216	0.52191	0.70676	0.83551
14	0.47549	0.60571	0.77606	0.88370
16	0.55329	0.67894	0.83088	0.91846
18	0.62518	0.74224	0.87374	0.94328
20	0.68973	0.79571	0.90670	0.96083
25	0.81546	0.89062	0.95776	0.98489
30	0.89672	0.94461	0.98173	0.99436
35	0.94510	0.97324	0.99239	0.99795
40	0.97207	0.98757	0.99693	0.99927
45	0.98633	0.99442	0.99879	0.99975
50	0.99353	0.99757	0.99954	0.99991
60	0.99868	0.99958	0.99994	0.99999
70	0.99975	0.99993	0.99999	1.00000
80	0.99996	0.99999	1.00000	1
90	0.99999	1.00000	1	1
100	1.00000	1	1	1

TABLE IX

(P_n - CUMULATIVE R_n)'s FOR THE TWO-TAILED EXPONENTIAL CASE WITH $\lambda=2$ AND FOR VARIOUS VALUES OF 2α

n	α = 0.01	0.02	0.05	0.10
1	0.00003	0.00010	0.00063	0.00250
2	0.00003	0.00012	0.00073	0.00295
3	0.00003	0.00012	0.00076	0.00308
4	0.00003	0.00012	0.00077	0.00312
5	0.00003	0.00012	0.00077	0.00314
6	0.00003	0.00012	0.00077	0.00315
7	0.00003	0.00012	0.00077	0.00315

$$R_3 = (\lambda y_{12} - \lambda y_{11} + 1)e^{-\lambda y_{12}} + (\tfrac{1}{2}\lambda^2 y_{12}^2 - \tfrac{1}{2}\lambda^2 y_{13}^2 + \lambda^2 y_{11}y_{13} - \lambda^2 y_{11}y_{12} + \lambda y_{11} - \lambda y_{13} - 1)e^{-\lambda y_{13}}. \tag{3.21}$$

For $\lambda = 2$ and $\alpha = 0.05$,

$$P_3 = 0.3165956$$

and $$\sum_{i=1}^{3} R_i = 0.3135211.$$

(3.14) - (3.17) and (iii) may be used to facilitate the computations of the P_n's and R_n's for any given α.

Power of the Method in the One-tailed Exponential Cases

Right Tail.

$$X_1, \ldots, X_n \sim \text{i.i.d. Exp}(\lambda),$$

and $$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling: sampling stops (with the rejection of H_0: $\lambda = 1$) the first time

$$s_n > \tfrac{1}{2}\chi_\alpha^2\,(2n).$$

(H_0 is to be tested against H_A: $\lambda < 1$.)

Let $$g_1(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

and $$g_n(x) = \lambda e^{-\lambda x}\int_0^{\min\{x, y_{n-1}\}} g_{n-1}(u)e^{\lambda u}.du, \quad x \geq 0,$$

where $y_{n-1} \equiv \frac{1}{2}\chi^2_\alpha\ (2(n-1))$ and $n = 2, \ldots$.

Defining

$$h_1(x) = 1, \quad x \geq 0,$$

and $$h_n(x) = \int_0^{\min\{x, y_{n-1}\}} h_{n-1}(u).du, \quad x \geq 0, \quad n = 2, \ldots,$$

then, analogous to (2.20),

$$\begin{aligned} P_n &\equiv P(M \leq n), \quad M \text{ as previously} \\ &= 1 - \lambda^n \int_0^{y_n} h_n(x) e^{-\lambda x}.dx, \quad n = 1, \ldots, \end{aligned}$$

and $$Q_n = \lambda^n \int_{y_n}^{\infty} h_n(x) e^{-\lambda x}.dx \quad \text{where}$$

Q_n denotes the probability of absorption in the boundary <u>at</u> the nth observation, $n = 1, \ldots$. Also, analogous to (2.21),

$$Q_n = \lambda^n e^{-\lambda y_n} \int_0^{y_{n-1}} h_{n-1}(u).du, \quad n = 2, \ldots .$$

Note that $\sum_{i=1}^{n} Q_i = P_n, \quad n = 1, \ldots$. (3.22)

<u>Left Tail</u>.

$$X_1, \ldots, X_n \sim \text{i.i.d. Exp}(\lambda),$$

and $$s_n \equiv \sum_{i=1}^{n} x_i$$

is used to decide whether to continue sampling, sampling stopping (with the rejection of H_0: $\lambda = 1$) the first time

$$s_n < \tfrac{1}{2}\chi^2_{1-\alpha}(2n).$$

(H_0 is being tested against H_A: $\lambda > 1$.)

Defining $h_1(x) = 1, \quad x \geq 0,$

and $$h_n(x) = \int_{y_{n-1}}^{x} h_{n-1}(u).du, \quad x \geq y_{n-1},$$

where $y_{n-1} \equiv \tfrac{1}{2}\chi^2_{1-\alpha}(2n-2)$ and $n = 2, \ldots ,$

then, analogous to (2.28),

$$P_n \equiv P(M \leq n), \quad M \text{ as usual}$$

$$= 1 - \lambda^n \int_{y_n}^{\infty} h_n(x)e^{-\lambda x}.dx, \quad n = 1, \ldots,$$

and $$R_n = \lambda^n \int_0^{y_n} h_n(x)e^{-\lambda x}.dx$$

is the probability of absorption in the boundary $\underline{at}$ the nth observation, $n = 1, \ldots$. Also, analogous to (2.29) and (2.30) respectively,

$$R_n = \lambda^n \int_{y_{n-1}}^{y_n} h_{n-1}(u)(e^{-\lambda u} - e^{-\lambda y_n}).du, \quad n = 2, \ldots,$$

and $$R_n = \lambda^n \int_{y_{n-1}}^{y_n} h_n(x)e^{-\lambda x}.dx, \quad n = 1, \ldots, \quad \text{where } y_0 \equiv 0.$$

Note that $\sum_{i=1}^{n} R_i = P_n, \quad n = 1, \ldots$.

CHAPTER IV

A PHILOSOPHICAL DISCUSSION ON THE RATIONALE OF METHODS OF SEQUENTIAL SAMPLING AND ANALYSIS

It is natural to question whether the criterion that has been used for determining the "significance" of results is legitimate. What is it that is rational or so special about the frequency characteristics that they should be chosen as the mode of inference rather than other possible methods? For example, "significance level" itself is not a well-defined entity (7). Easterling (18) in an excellent article addressed to "Reliability engineers, statisticians, and Bayesians" discusses much that is both pertinent and very mundane:

> It is really not appropriate to lump all non-Bayesian approaches to statistical inference under one heading. However, since the expression "classical statistics" has some currency, though no precise definition, we shall let it stand as a heading. ...
>
> The test of significance is a concept due to R. A. Fisher...he developed the test of significance to answer the question, "to what extent are the data consonant with a given hypothesis?" ... To answer this he proposed the statistic: the relative frequency in repetitions from a hypothetical population in which results as extreme or more so as that observed are obtained, where by more extreme we mean those hypothetical results which support the alternative to the hypothesis being tested more than they support the hypothesis ... It may help to think of...repeated experimentation, but this interpretation is not necessary and often untenable. ...

> Another objection is against the use of tail areas. Kempthorne [29 here] supports this measure by describing the significance test as a measure of the distance x is from the hypothetical data which are generated by $f(X; \theta_0)$. ... The reason the significance test is used is because it has certain desirable operating characteristics.
>
> With this basic tenet, that operating characteristics are informative and pertinent, I am willing to consider any statistic regardless of its origin. I see no need to adopt any one "optimality" criterion, such as unbiasedness, maximum likelihood, or the best Bayes decision rule to derive acceptable statistics. ...
>
> I can sympathize with the effort to bring a consistent logic to statistical practice. But I do not feel inadequate because of the absence of this (pp. 190-192).

Anscombe (2) has asserted that "All risk of error is avoided if the method of analysis uses the observations only in the form of their likelihood function, since the likelihood function (given the observations) is independent of the sampling rule" (page 100).

McPherson and Armitage (41) have perhaps the most relevant comments:

> Analyses of data by likelihood functions or posterior probabilities are completely unaffected by stopping rules; tail-area significance tests, by contrast, are highly sensitive to the stopping rule. However, the probability of achieving a particular result measured by likelihoods or posterior probabilities is affected by the number of times the data are examined. Certain applications of likelihoods or posterior probabilities lead to the same stopping rules as would repeated significance tests at a fixed nominal level. For instance, if the ratio of the likelihood of the hypothesis to the maximum likelihood is tested after each observation in $N(\mu, 1)$ variates, a reasonable stopping rule is: stop iff $L_0/L_{Max} \leq$ some constant r. This is equivalent to repeated significance tests at a two-sided level $2\alpha^*$, where α^* is given by
>
> $$\Phi\left[\sqrt{\{2 \log_e(1/r)\}}\right] = 1 - \alpha^*. \qquad (3)$$
>
> If, similarly, for $N(\mu, 1)$ variates we postulate that the prior distribution of μ is $N(0,\sigma_0^2)$, and measure the posterior probabilities

that μ is greater than or less than zero at each observation, we might stop iff

$$\int_{-\infty}^{0} \pi(\mu/s_n) \leq \lambda$$

or (4)

$$\int_{0}^{\infty} \pi(\mu/s_n) \leq \lambda \quad ,$$

where $\pi(\mu/s_n)$ is the posterior density of μ. This leads to the stopping rule: stop iff

$$|s_n| \geq k_2\sqrt{(n + \sigma_0^{-2})} \qquad (5)$$

where $\Phi(k_2) = \frac{1}{2}-\lambda$. Where the prior distribution is uniform, $\sigma_0^{-2} = 0$ and the stopping rule is equivalent to repeated significance tests at a two-sided level of 2λ.

Hence ... repeated significance tests ... provide a basis for sequential analysis which [is] capable of interpretation from a frequentist, likelihood or Bayesian approach (page 20).

Thus, the frequentist mode of inference used in at least one section of each of Chapters II and III (namely the two-tailed normal case) is equivalent to both a likelihood ratio approach and a Bayesian approach (with a vague prior). The same is also true of the left-tailed and right-tailed normal cases; i.e. the frequentist mode of inference used in the one-tailed normal case for testing that $N(\mu, 1)$ variates come from a population whose mean is zero (H_0: $\mu = 0$) against either that the population mean is negative (H_A: $\mu < 0$) or positive (H_A: $\mu > 0$) is equivalent to both a likelihood approach and a Bayesian approach (with a vague prior). After a digression into these approaches this critically important topic will be reintroduced in Chapter V.

Some relevant comments on likelihood, likelihood ratio and likelihood principle are now given. This section will then be followed by a discussion on Bayesian techniques. These two positions will be seen to be intimately connected.

Likelihood Approach

There seem to be as many versions of the so-called "likelihood principle" as there are authors who write on it! (c.f. (16), (30) and (45).) As Kempthorne and Folks (30, page 295) have it:

> This [the likelihood principle] has not bee stated tightly but appears to be as follows. 'To form opinions about parameter values from data, the only inferential content of the data is given by the realized likelihood function.'

L. J. Savage (45, pages 184, 185) was more commital:

> From the Bayesian position heretofore scattered ideas take on new unity and comprehensibility.
>
> One of the most obvious, ubiquitous and valuable consequences of the Bayesian position is what I call the likelihood principle. This principle was, so far as I know, first advocated by George Barnard [8 here].
>
> ... 'the likelihood function, long known to be a minimal sufficient statistic, is much more than merely a sufficient statistic, for given the likelihood function in which an experiment has resulted, <u>everything</u> else about the experiment - what its plan was, what different data might have resulted from it, the conditional distributions of statistics under given parameter values, and so on - is irrelevant.'
>
> ... The likelihood ... retains its import even if the experiment terminated merely when the experimenter happened to get tired or run out of time - always under the proviso that the individual trials are independent. ...
>
> This same function even persists if the experimenter quits only when he believes he has enough data to convince others of his own opinion. This leads to the moral that optional stopping ... is no sin, but that traditional methods of judging data in terms of significance level cannot safely be interpreted without regard to other information.

Cornfield (16) mentions <u>preserving</u> (which should be determining)

> <u>the</u> critical level, i.e. <u>the</u> lowest significance level at which the hypothesis can be rejected for given data. ... the critical level provides <u>an</u> appropriate measure of the amount

> of evidence [?] in the data for or against the hypothesis. ... The critical level is thus regarded as a universal yardstick (page 18).

(The emphasis has been added here and in the following.) Unfortunately the usage <u>here</u> of the terms "critical level" and (prechosen) "significance level" is as given by Lehmann (34, pages 61 and 62), which is less commonly accepted than reversing the roles played by these terms. Cornfield later confuses the two! He then gives what he references as the α-postulate: "All hypotheses rejected at <u>the</u> same critical level have <u>equal amounts of evidence</u> [?] against them." He admits that he has never seen nor heard this postulate explicitly stated, nor can he name any <u>statistician</u> who believes it, but asserts that <u>he</u> believes that sequential analysis can be defended if and only if "something like" the α-postulate is true!

Cornfield then attempts to demolish his own argument! Three examples are proposed and each is claimed to refute Cornfield's α-postulate. Curiously not one succeeds! The third example is:

> (c) D. R. Cox [17 here] has constructed an example which suggests that the most powerful test of the hypothesis that a mean is zero against a particular alternative will sometimes reject the null hypothesis when the observed mean is zero (page 19).

The quoted reference has no such fabrication! Even if it did there are much simpler contrivances which illustrate the point Cornfield (irrelevantly) tries to make: for random variable $X \sim N(\theta, 1)$, consider the uniformly most powerful test of H_0: $\theta = 0$ against H_A: $\theta > 0$ using $\alpha = 0.6$.

"But if one is willing to be guided by the α-postulate...why should he be any more willing to accept it when analyzing sequential trials?" Categorically, one need not accept it in sequential methods but <u>may</u>

appear to do so only in the name of mathematical convenience - only for the sake of standardizing a procedure!

Cornfield then turns to his second line of argument

> - which is that there is a reasonable alternative explication of the idea of inference and one which leads to the rejection of sequential analysis. This explication is provided by the likelihood principle - which states that all observations leading to the same likelihood function should lead to the same conclusion (page 20).

The likelihood functions of the binomial and negative binomial are then discussed. To fill in omitted details: consider n (or N) independent dichotomous trials, each with constant non-zero probability p of a "success", leading to r (or R) successes. If n is a pre-specified positive integer then R is a random variable whose distribution is given by

$$P(R = r) = \begin{cases} \binom{n}{r} \; p^r \; (1-p)^{n-r}, & r = 0, 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

If r is a pre-specified positive integer, i.e. continue random sampling until the r^{th} success occurs then stop, then N is a random variable whose distribution is given by

$$P(N = n) = \begin{cases} \binom{n-1}{r-1} \; p^r \; (1-p)^{n-r}, & n = r, r+1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

The factors which depend on parameter p, namely $p^r(1-p)^{n-r}$ <u>in each case</u>, is regarded as the likelihood function. The argument continues that since both distributions yield the one likelihood function, if one accepts the likelihood principle one "must come to the same common conclusion about p, despite the use of quite different stopping rules." Using "some different inferential principle, say that of unbiased estimation, however, the first investigator would have estimated p as as r/n and the second as $(r-1)/(n-1)$." No mention is made of the restriction $r > 1$ necessary in the latter case. Nor is mention made as to why unbiasedness should be used as the hallowed "inferential principle": it is well-known that likelihood techniques and unbiasedness are at variance - for random variable $X \sim N(\mu, \sigma^2)$, both μ and σ^2 unknown, the maximum likelihood estimate of σ^2 is biased. Cornfield concludes that "if one accepts the likelihood principle one must reject sequential analysis" (page 20).

Now the situation will be re-analyzed, this time without slipping over the crucial stepwise meaning of the symbols, for it is within this new framework that the rebuttal to the argument will be seen to lie - it will be seen that the 'old' argument became lost in the unquestioned mathematical symbolism!

What is meant by the term 'likelihood function'? For present purposes, X being a random variable whose probability <u>mass</u> function will be denoted by $p(x; p)$, single parameter $p \in (0,1)$, and $x_1, \ldots, x_n$ being a random sample from this distribution, then the likelihood function is given by

$$L(p/\underset{\sim}{x}) \equiv \prod_{i=1}^{n} p(x_i; p).$$

Thus in the binomial case there are purportedly n independent observations $r_1, \ldots, r_n$ from

$$P(R = r_i) = \begin{cases} p^{r_i} (1-p)^{1-r_i}, & r_i = 0, 1, \\ 0 & \text{otherwise}, \end{cases} \qquad i = 1, \ldots, n,$$

so the likelihood function here is given by

$$L_B(p/\underset{\sim}{r}) \equiv p^r(1-p)^{n-r} \quad \text{where} \quad r \equiv \sum_{i=1}^{n} r_i \quad .$$

In the negative binomial case there are r' independent observations $n_1, \ldots, n_{r'}$, from

$$P(N = n_j) = \begin{cases} p(1-p)^{n_j - 1}, & n_j = 1, \ldots, \\ 0 \quad \text{otherwise}, & \end{cases} \qquad j = 1, \ldots, r',$$

i.e. n_j is the number of trials between the $(j-1)^{th}$ and j^{th} successes not counting the trial on which the $(j-1)^{th}$ success occurred but counting the trial on which the j^{th} success occurred, so the likelihood function here is given by

$$L_N(p/\underset{\sim}{n}) \equiv p^{r'} (1-p)^{n'-r'} \quad \text{where} \quad n' = \sum_{j=1}^{r'} n_j .$$

Now $$L_N(p/\underset{\sim}{r}) \equiv L_B(p/\underset{\sim}{n}) , \quad p \in (0,1)$$

$$\iff \quad r' = r$$
$$\text{and} \quad n' = n,$$

i.e. the two likelihood functions are identical if and only if

(i) the number of successes in the binomial case is equal to the pre-specified number of successes in the negative binomial case,

(ii) the number of trials required in the negative binomial case is equal to the pre-specified number of trials in the binomial case, and

(iii) the last trial resulted in a success for certain (and not the first success at that): this is taken into consideration in the negative binomial case - it is a pre-condition - but not in the binomial case.

Given that the experiment resulted in identical likelihood functions then the last trial of the binomial experiment was non-random (since a success certainly occurred on this trial). Then this observation, being non-random whereas those preceding it were random, should be discarded - it contains no information (in any sense) about p. Thus in the binomial case the experiment should be considered as consisting of n-1 independent trials resulting in r-1 successes, and Cornfield's 'contradiction', even based on unbiasedness, is resolved.

Finally, D. R. Cox (17, pages 363-366) has given his views:

> In the problem without nuisance parameters, it is known that methods of inference ... that use only observed values of the likelihood ratios, and not tail areas, avoid the difficulties ... since the likelihood ratio is the same whether we argue conditionally or not.
>
> [Writing on the Bayesian approach] An important advantage of this approach is that it ensures independence from the sampling rule ... . [See Anscombe (1).]

Bayesian Approach

For present purposes it suffices to characterize the Bayesian viewpoint in the following way:

X is a random variable with <u>density</u> $f(x;\ \theta)$ where the 'parameter of interest' $\theta \in \Omega$, the parameter set (or space); θ itself is now considered as a random variable Θ with prior <u>density</u> denoted by $\pi_0(\theta)$. One may think of π_0 as being, in <u>some</u> intuitive sense, the "best description of the distribution of θ available in the absence any (further) data." A random sample $X_1, \ldots, X_n$ is then taken from $f(x;\ \theta)$, which should now be written $f(x/\theta)$, and 'summarized' by statistic $Y = Y(X)$, sufficient for θ. Furthermore, suppose Y has density $g(y/\theta)$ (this being essentially the likelihood

$L(\theta/\underset{\sim}{x})$); then the posterior density of Θ (with motivation via Bayes's theorem for absolutely continuous random variables) is defined to be

$$\pi_1(\theta/y) \equiv \frac{g(y/\theta)\pi_0(\theta)}{\int_\Omega g(y/\theta)\pi_0(\theta).d\theta} \qquad (4.1)$$

(assuming the right-hand side here exists).

Hopefully π_1 is, in some intuitive sense, the "best description of the distribution of Θ available after the data has been taken." The posterior density π_1 of Θ is then the inference base for θ. The chosen prior distribution and the data have been merged via Bayes's theorem to yield a posterior distribution: one may think of the posterior as being, in some intuitive sense, how the data has modified the chosen prior. Notationally Θ_0 will represent the prior random variable and Θ_1 will denote the posterior random variable.

To exemplify some points consider

$$X \sim N(\theta, \sigma^2), \quad \theta \in R \equiv (-\infty, \infty) = \Omega,$$

$$\sigma^2 \text{ known (positive)},$$

$$\Theta_0 \sim N(\mu_0, \sigma_0^2), \quad \mu_0 \text{ known (real)}, \quad \sigma_0^2 \text{ known (positive)},$$

and $X_1, \ldots, X_n$ is a random sample of X's

so $Y \equiv \bar{X}$ here is sufficient for θ

$$\sim N(\theta, \frac{\sigma^2}{n});$$

then
$$\Theta_1 \sim N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}, \ \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \right). \qquad (4.2)$$

For $n = 1$,

$$\Theta_1 \sim N\left(\frac{\frac{x_1}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}} \equiv \mu_1, \ \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}} \equiv \sigma_1^2\right).$$

Thus, having randomly sampled a single observation x_1 the (first) posterior distribution at this stage is as given.

Now follow an "empirical Bayes" procedure: use this distribution as the prior for a second randomly sampled single observation x_2 (independent of x_1). The second posterior random variable

$$\Theta_2 \sim N\left(\frac{\frac{x_2}{\sigma^2} + \frac{\mu_1}{\sigma_1^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}}, \ \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}}\right),$$

i.e.

$$N\left(\frac{\frac{x_1+x_2}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{2}{\sigma^2} + \frac{1}{\sigma_0^2}}, \ \frac{1}{\frac{2}{\sigma^2} + \frac{1}{\sigma_0^2}}\right).$$

This process can be repeated ad infinitum and on the p^{th} repetition ($p = 0, 1, \ldots$) the p^{th} posterior random variable

$$\Theta_p \sim N\left(\frac{\frac{\sum_{i=1}^{p} x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{p}{\sigma^2} + \frac{1}{\sigma_0^2}}, \ \frac{1}{\frac{p}{\sigma^2} + \frac{1}{\sigma_0^2}}\right). \tag{4.3}$$

Taking $\sigma^2 = 1$

$= \sigma_0^2$,

the crux of the so-called Bayesian controversy - or rather it is the beginning thereof, but by no means the end! If one can justify the choice of prior in some meaningful way that was "acceptable" (as opposed to completely contrived) application in the real world then apply Bayes's theorem: the use of Bayes's theorem does not make one a Bayesian and it is well-known that, as Easterling (18) puts it, "one must bear in mind that posterior probability statements are conditional on the prior." D. R. Cox (17) uses the terms "... an agreed prior ...", " ... conventional form of prior ... " and qualifies one statement with "when the choice of prior is difficult." In sharp contrast Lindley (37, page 421) has objected

> to the statement, repeatedly made, that a prior is unknown. This is ridiculous, a prior is a statement of one's knowledge and ~~modern work~~ demonstrates that it is always known: by judicious questioning it can be found.

Easterling (18, page 189) has made the very pertinent point that

> ... it is critical that the results of the experiment stand alone so that they can be added to the store of knowledge and so that others can draw their own conclusions. Bayes' Theorem merges these two items, sometimes inextricably.

Barnard (6, page 194) had previously noted this, though not as forcefully:

> The main quarrel I have with the subjective Bayesian approach is this, that I fear that it does not always make clear to the client or consumer how much of the message presented to him in the form of a posterior distribution really comes from the data and how much from the assumption involved in the prior distribution.

Bayes's theorem merges the chosen prior and the observed data in a rigid manner - it does not allow for any weighting of the data with respect to the chosen prior. This objection may be overcome; the rationale and motivation for the method employed will be given first. The argument is entirely verbal: it rests completely on intuitive appeal.

Recall the verbalizations that

(i) the prior may be thought of as the best description of the distribution of random variable Θ available before the data $\underset{\sim}{x}$ is observed, and

(ii) hopefully the posterior is the best description of the distribution of Θ available after the data has been taken.

The posterior is to be considered "superior" to the prior for the purpose of inference about θ. (Otherwise the prior would be used for this purpose!) Hence if one knew the posterior before randomly sampling one would surely use this distribution as the prior, thus obtaining an even better posterior than the "original" posterior. The data are more heavily weighted than originally by Bayes's theorem! Notationally Θ_2 will represent this second posterior random variable and π_2 will represent its density.

To illustrate this procedure, by analogy with (4.1),

$$\pi_2(\theta/y) \equiv \frac{g(y/\theta)\pi_1(\theta)}{\int_\Omega g(y/\theta)\pi_1(\theta).d\theta}, \quad \text{where} \quad \pi_1(\theta) \text{ is to be}$$

interpreted as $\pi_1(\theta/y)$ given by (4.1)

$$= \frac{g(y/\theta)\ \dfrac{g(y/\theta)\pi_0(\theta)}{\int_\Omega g(y/\theta)\pi_0(\theta).d\theta}}{\int_\Omega g(y/\theta)\ \dfrac{g(y/\theta)\pi_0(\theta)}{\int_\Omega g(y/\theta)\pi_0(\theta).d\theta}.d\theta}$$

$$= \frac{g^2(y/\theta)\,\pi_0(\theta)}{\int_\Omega g^2(y/\theta)\,\pi_0(\theta).d\theta}$$ (assuming the right-hand side here exists).

This procedure can be repeated sequentially: more and more weight is put on the data (with respect to the original chosen prior). With a natural and self-explanatory extension of notation, for $p = 3, 4, \ldots,$

$$\pi_p(\theta/y) \equiv \frac{g(y/\theta)\,\pi_{p-1}(\theta)}{\int_\Omega g(y/\theta)\,\pi_{p-1}(\theta).d\theta},$$ where $\pi_{p-1}(\theta)$ is to be interpreted as $\pi_{p-1}(\theta/y)$ from the previous step (assuming existence)

$$= \frac{g^p(y/\theta)\,\pi_0(\theta)}{\int_\Omega g^p(y/\theta)\,\pi_0(\theta).d\theta}$$

(assuming existence). This equation holds for all $p = 1, \ldots$. It may also be interpreted as holding for $p = 0$ providing π_0 is "normed" to unity, i.e. integrates to unity on the real line (which can be taken for granted without loss of generality providing $\int_\Omega \pi_0(\theta).d\theta < \infty$, in which case π_0 is called "proper").

This procedure shares some properties with the empirical Bayes technique, but the two are quite distinct. For one, the empirical Bayes technique requires that a random sample be taken between calculation of posteriors and this is not the case with the above technique.

Without enquiring further what this procedure does and means and why it is done here, one immediately asks a question that is begged:

"Does Θ_p have a limit as p tends to infinity?" (i.e. "What is Θ_∞?") The following examples provide some answers:

<u>Example 4.1:</u> If $X \sim N(\theta, \sigma^2)$, $\theta \in R = \Omega$, σ^2 known (non-negative),

$$\Theta_0 \sim N(\mu_0, \sigma_0^2), \quad \mu_0 \text{ known}, \quad \sigma_0^2 \text{ known (positive)},$$

and $X_1, \ldots, X_n$ are independent X's so $Y \equiv \bar{X}$ here is sufficient for θ $\sim N(\theta, \frac{\sigma^2}{n})$,

then
$$\Theta_p \sim N\left(\frac{\frac{np\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{np}{\sigma^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{np}{\sigma^2} + \frac{1}{\sigma_0^2}}\right), \quad p = 0, 1, \ldots,$$

so Θ_∞ is degenerate at $y \equiv \bar{x}$ $\forall\ \mu_0 \in R$. This is true $\forall\ \sigma^2 > 0$, $\forall\ \sigma_0^2 > 0$ and $\forall\ n = 1, \ldots$. In the limit the weight on the data is so heavy with respect to the chosen prior as to wash out the effect of the prior: according to Easterling (18), the

> coincidence of Bayesian and classical results brings to mind one rationale that some advance as support for the Bayesian approach, which is that if one has enough data, the effect of the prior is washed out (page 188).

For a (proper) vague prior take $\lim_{\sigma_0^2 \to \infty}$; then

$$\Theta_p \sim N(\bar{x}, \frac{\sigma^2}{np}), \quad p = 1, \ldots .$$

Hence not only is X(hence $\bar{X}$) unbiased for θ here, but also, under this vague prior, Θ_p is unbiased for $\bar{x}$ (or just x), $n = 1, \ldots,$ $p = 1, \ldots$.

<u>Example 4.2:</u> If $X \sim Exp(\theta)$, $\theta > 0$ (i.e. $\Omega = (0, \infty)$),

$$\Theta_0 \sim Exp(\lambda), \quad \lambda \text{ known (positive)},$$

and $X_1, \ldots, X_n$ is a random sample of X's so $Y \equiv \sum_{i=1}^{n} x_i$ here is sufficient for θ

$$\sim Ga(\theta, n),$$

i.e. $$g(y/\theta) = \begin{cases} \frac{1}{(n-1)!}\, \theta^n\, y^{n-1}\, e^{-\theta y}, & y > 0, \\ 0 & \text{otherwise}, \end{cases}$$

then $\forall\ \theta > 0$,

$$\pi_p(\theta/y) = \frac{\theta^{np}\, e^{-py\theta}\, e^{-\lambda\theta}}{\int_0^\infty \theta^{np}\, e^{-py\theta}\, e^{-\lambda\theta}\ .d\theta},$$

so $\Theta_p \sim Ga(np\bar{x} + \lambda,\ np+1)$, $p = 0, 1, \ldots$.

The characteristic function of Θ_p is then

$$\phi_p(t) \equiv \left(1 - \frac{it}{py+\lambda}\right)^{-(np+1)}$$

$$\rightarrow e^{\lim_{p\to\infty} \frac{np+1}{py+\lambda} it} \quad \text{as } p \to \infty$$

$$\equiv e^{\frac{it}{\bar{x}}}$$

so by the Levy-Cramer theorem (Fisz (23), for example) Θ_∞ is degenerate at $\frac{1}{\bar{x}}$. This is true $\forall\ \lambda > 0$ and $\forall\ n = 1, \ldots$. Again the increasingly heavy weight on the data has washed out the effect of the prior chosen here!

For a (proper) vague prior take $\lim_{\lambda \downarrow 0}$; then $\Theta_p \sim Ga(np\bar{x},\ np+1)$, $p = 0, 1, \ldots$. Thus not only is X (hence $\bar{X}$) unbiased for $\frac{1}{\theta}$ here, but also, under this vague prior, $\frac{1}{\Theta_p}$ is unbiased for $\bar{x}$ (or just x), $p = 1, \ldots$, since for $p = 1, \ldots$,

$$E\left(\frac{1}{\Theta_p}\right) = \frac{(py)^{np+1}}{(np)!} \int_0^\infty \frac{1}{\Theta}\, \Theta^{np}\, e^{-py\Theta}\ .d\Theta$$

$$= \bar{x}, \qquad n = 1, \ldots .$$

In contrast, $\frac{1-\frac{1}{n}}{\bar{X}}$ is unbiased for θ, $n = 2, 3, \ldots$ $(n \neq 1)$, while under the given vague prior $\frac{\Theta_p}{1+\frac{1}{np}}$ is unbiased for $\frac{1}{\bar{x}}$, $n, p = 1, \ldots$. (The first estimate affords a situation in which at least two population units would be sampled at a time. The second estimate, in considering $E(\Theta_p)$, essentially utilizes the squared-error loss function.) It is also of academic interest to note that both these estimates have rather obtuse analogues in normal distribution theory:

Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, both μ and σ^2 unknown $(\sigma^2 > 0)$; then $\frac{1-\frac{1}{n}}{\left(\frac{1}{\sigma^2}\right)}$ is unbiased for $\overline{(x-\bar{x})^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ while $\frac{1}{1+\frac{1}{n}}\overline{(x-\bar{x})^2}$ is the minimum mean-square error estimate of $\frac{1}{\left(\frac{1}{\sigma^2}\right)}$.

A natural extension of this weighting method leads to an interesting conclusion. By holding the philosophy that the observed data $\underset{\sim}{x}$ in some sense reflects something informative about the (realized or present) value θ of random variable Θ, and supposing

(i) the prior to be not just the best description of the distribution of Θ (before the observations $\underset{\sim}{x}$ are taken) but the <u>true</u> distribution of Θ, and

(ii) $\underset{\sim}{x}$ is, <u>as a random sample</u>, representative of the whole population (of which $f(x/\theta)$ is the density), i.e. assuming the data $\underset{\sim}{x}$ are "obliging" for the purpose of inference about θ, then

the posterior returned from merging the data and the prior via Bayes's theorem may reasonably be expected to be just the prior; i.e., dropping the subscripts on the prior and posterior densities π_0 and π_1, the following functional equation is of interest:

$$\pi(\theta) = \frac{g(y/\theta)\pi(\theta)}{\int_\Omega g(y/\theta)\pi(\theta).d\theta} .$$

For given $g(y/\theta)$ this equation is to be solved for $\pi(\theta)$. Hence for almost all y,

$$\int_\Omega g(y/\theta)\pi(\theta).d\theta = g(y/\theta)$$ almost everywhere with respect to the probability measure π on Θ.

Now the left-hand side of this equation is independent of θ, so $g(y/\theta)$ is independent of θ! This seems to contradict the philosophy that $\underset{\sim}{x}$ reflects something about the value θ of Θ. Then surely the only conclusion is that the posterior must be different from the prior (on some subset of Ω of non-zero prior and posterior measure): the data must modify the prior - either for better or worse!

Jeffreys (27) has rationalized a vague prior for binomial parameter p:

$$\pi_0(p) = \begin{cases} \frac{1}{p(1-p)} , & 0 < p < 1, \\ 0 & \text{otherwise}. \end{cases}$$

Note that $\forall\ \varepsilon \in (0, \frac{1}{2})$,

$$\int_\varepsilon^{1-\varepsilon} \pi_0(p).dp < \infty$$

but that

$$\lim_{\varepsilon \downarrow 0} \int_\varepsilon^{1-\varepsilon} \pi_0(p).dp = \infty .$$

This much-discussed prior is unusual for its properties - tending to put infinitely more prior weight in the interval $(0, \varepsilon)$ and again in the interval $(1-\varepsilon, 1)$ than in the in-between interval $(\varepsilon, 1-\varepsilon)$. In discussing such "improper" priors Hacking (26, page 204) writes:

> If we have an unknown parameter which can range anywhere from 0 to ∞, we are usually told to assume that the prior probability of the logarithm of the parameter is uniformly distributed. So we assent to probabilities that do not sum to any finite quantity. We substitute these in a formula, use some other data, and get probabilities that sum to 1. What is going on here? It looks like magic ... .

According to Perks (43, pages 55 - 57), Jeffreys modified this prior to

$$\pi_0(p) = \begin{cases} \dfrac{1}{\sqrt{p(1-p)}}, & 0 < p < 1, \\ 0 & \text{otherwise,} \end{cases}$$

so that $\int_0^1 \pi_0(p).dp = \pi.$

Novick (43, pages 61 - 64), Lindley (43, pages 57 - 58) and I. J. Good (43, pages 59 - 61) have provided further discussion on this.

To round out this discussion on the Bayesian approach both "camps" will have their say:

> Indeed the whole Bayesian computation is trivally easy providing that one slips over the question of what the meaning of the result is ... I am opposed to the type of thinking ... that the best approach to data interpretation is to feed the data through the Bayesian process with a prior that is arbitrary (or perhaps has mathematical convenience).
>
> \- Kempthorne (15, pages 648, 653)

> ... prior distributions are often specified and used when they are not describing a real random process nor deduced in a logical manner to describe a certain state of knowledge. The introduction of such an element into the inference seems to us quite unscientific. We do not agree that the purpose of a scientific investigation and the subsequent statistical analysis is to quantify personal belief and so that justification for the use of such priors is not acceptable to us.
>
> \- Kalbfleisch and Sprott (28, page 206)

Box and Tiao (12, page 9-10) on "The Role of Bayesian Analysis":

> Because this system of inference may be readily applied to any probability model, much less attention need be given to the <u>mathematical</u> convenience of the models considered and more to scientific merit. ...
>
> It is, we believe, equally unhelpful for enthusiasts to ... claim that Bayesian analysis can do everything, as it is for its detractors to ... assert that it can do nothing.

> I believe that the lesson that we must learn is that there is no single theory entirely free from deficiencies. We have to be willing to learn about the advantages and disadvantages of all concepts used in inference about certainty. We owe a great deal to the Bayesian school of thought but we do object to a dogma in which this philosophy is worshipped as the infallible and completely virtuous solution of the decision maker.
>
> \- Hartley (15, page 647)

From Geisser (15, page 645) on Bayesians: "'Ye shall know them by their posteriors.'"

CHAPTER V

THE EXPONENTIAL CASES REVISITED

As noted early in Chapter IV the frequentist mode of inference used in the normal cases in Chapters II and III is equivalent to both a likelihood ratio and a Bayesian approach (with a vague prior). These approaches will now be investigated in relation to the two-tailed exponential case. The one-tailed exponential cases are simplifications of this case.

A Likelihood-Frequentist Approach

For $X \sim \text{Exp}(\lambda), \quad \lambda > 0,$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

so that $\forall\, x_i > 0, \quad i = 1, \ldots, n,$

$$L(\lambda/\underset{\sim}{x}) \equiv \lambda^n e^{-\lambda y}, \quad y \equiv \sum_{i=1}^{n} x_i, \quad n = 1, \ldots .$$

Suppose one is interested in testing $H_0\colon\ \lambda = \lambda_0$ against $H_A\colon\ \lambda \neq \lambda_0$.

$$\frac{L_0}{L_{Max}} = (\lambda_0 \bar{x})^n e^{-n(\lambda_0 \bar{x} - 1)}, \quad \text{where} \quad \bar{x} \equiv \bar{x}(n)$$

$$\equiv \frac{1}{n} \sum_{i=1}^{n} x_i,$$

$$n = 1, \ldots .$$

Let $r \equiv r(2\alpha, n)$

$\varepsilon\ (0, 1)$ and such that

$\frac{L_0}{L_{Max}} < r$ defines a critical region of nominal size 2α ($\alpha \ \varepsilon\ (0,\frac{1}{2})$) for testing H_0 against H_A using a fixed-sample-size procedure;

then a rational and reasonable stopping rule is: sampling stops (with the rejection of H_0) the first time

$$n(1 - \lambda_0\bar{x} + \ln(\lambda_0\bar{x})) < \ln r,$$

i.e.

$$w - \ln w > 1 - \frac{1}{n}\ln r, \quad w \equiv \lambda_0\bar{x}.$$

(The appearance of the intuitive "reasonableness" of this stopping rule is to some extent analogous to the apparent "reasonableness" of consideration of highest posterior density regions of Bayesian methods.)

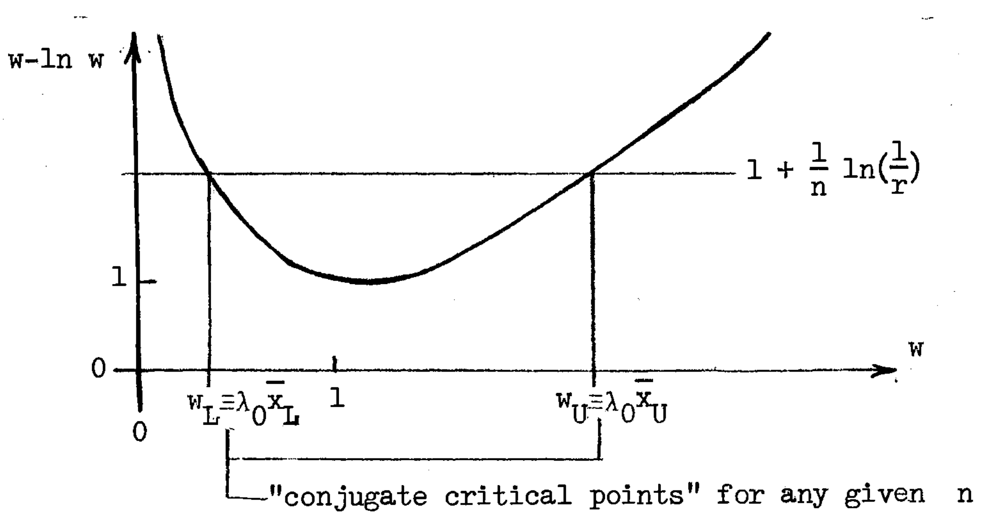

Figure 1. Graph of $w - \ln w$ Against w for Any Given n

Without loss of generality take $\lambda_0 = 1$ (otherwise take $\frac{\lambda}{\lambda_0}$ in place of λ and $\lambda_0 x$ in place of X), so under H_0,

$$2X \sim \chi^2(2) \equiv \mathrm{Exp}(\tfrac{1}{2});$$

then one is interested in solving

$$\bar{x} - \ln \bar{x} = 1 + \frac{1}{n} \ln \frac{1}{r}$$

$$\equiv c_n \text{ for } \bar{x}_L \text{ and } \bar{x}_U ,$$

i.e. $x^{(n)} - \ln x^{(n)} = c_n$ for $x_L^{(n)} < 1$ and $x_U^{(n)} > x_L^{(n)}$

subject to (since $2nX^{(n)} \sim \chi^2(2n)$)

$$F_{2n}(2nx_L^{(n)}) + 1 - F_{2n}(2nx_U^{(n)}) = 2\alpha, \text{ where } F_{2n} \text{ is}$$

the cumulative distribution function of a chi-squared random variable with $2n$ degrees of freedom,

i.e. $x_L^{(n)} - \ln x_L^{(n)} = x_U^{(n)} - \ln x_U^{(n)}$ (5.1)

subject to

$$F_{2n}(2nx_U^{(n)}) - F_{2n}(2nx_L^{(n)}) = 1 - 2\alpha. \qquad (5.2)$$

(5.1) and (5.2) are to be solved simultaneously for

$$nx_L^{(n)} < n \text{ and } nx_U^{(n)} > nx_L^{(n)} \quad \text{for any given } \alpha \in (0,\tfrac{1}{2})$$

and $\forall\, n = 1, \ldots, n_{Max}$: $nx_L^{(n)}$ and $nx_U^{(n)}$ will replace y_{1n} and y_{2n}, respectively, in the two-tailed exponential case of hypothesis testing at a nominal 2α level after each observation has been randomly sampled.

Results are given in Table X.

TABLE X

VALUES OF $nx_L^{(n)}$, $nx_U^{(n)}$, (i) $P(\chi^2(2n) < 2nx_L^{(n)})$,
(ii) $P(\chi^2(2n) > 2nx_U^{(n)})$, $r(2\alpha, n)$ AND
P_n FOR $2\alpha = 0.10$

n	$nx_L^{(n)}$	$nx_U^{(n)}$	(i)	(ii)	r(0.10, n)	P_n
1	0.083815	3.932145	0.08040	0.01960	0.20952	0.10
2	0.441327	5.479177	0.07296	0.02704	0.23141	0.16253
3	0.937295	6.946117	0.06914	0.03086	0.23993	0.20485
4	1.508663	8.355396	0.06673	0.03327	0.24440	0.23651
5	2.129108	9.723134	0.06505	0.03495	0.24714	0.26170
6	2.78479	11.0595				0.28256
7	3.46737	12.3712				0.30031
8	4.17137	13.6629				0.31575
9	4.89294	14.9379				0.32938
10	5.62928	16.1989				0.34158
15	9.47174	22.3483				0.38827
20	13.4934	28.3226				0.42091
30	21.8489	39.9630				0.46576
40	30.4607	57.3492				0.49653
50	39.2365	62.5721				0.51969

By comparing Tables II (with $2\alpha = 0.10$) and X it may be observed that up to $n = 5$ the P_n's of Table II are less than those of Table X, while for larger values of n the opposite is true. This may be suggesting that sequential testing based on not only frequency characteristics but also on the likelihood ratio is, for sufficiently large sample sizes n, more powerful than one based on frequency characteristics alone.

From these results, for

$$P_1 \equiv 2\alpha$$
$$= 0.10,$$

$$P_2 = 1 + e^{-y_{21}} + (y_{11} - y_{12} - 1)e^{-y_{12}} + (y_{21} - y_{11})e^{-y_{22}} \quad (2.11)$$

where $y_{1n} \equiv nx_L^{(n)}$

and $y_{2n} \equiv nx_U^{(n)}$, $n = 1, 2$

$$= 0.162532 \quad (\text{for } 2\alpha = 0.10)$$

and from (2.18),

$$P_3 = 0.204846.$$

A Pure Likelihood Ratio Approach

Fix $r = r(2\alpha, 1)$

$= 0.209515$ for $2\alpha = 0.10$

in the Likelihood-Frequentist Approach, so that

$$P_1 = 2\alpha;$$

then one is interested in solving

$$x^{(n)} - \ln x^{(n)} = 1 - \frac{1}{n}\ln r \qquad (5.3)$$

$$\text{for } nx_L^{(n)} < n \text{ and } nx_U^{(n)} > n$$

$$= 1 + \frac{1.56296}{n} \text{ for } 2\alpha = 0.10.$$

(5.3) is to be solved for $nx_L^{(n)} < n$ and $nx_U^{(n)} > n$ for any given $(0 < \alpha << 0.5)$ and $\forall\ n = 1, \ldots, n_{Max}$: $nx_L^{(n)}$ and $nx_U^{(n)}$ will again replace y_{1n} and y_{2n}, respectively, in the two-tailed exponential case of hypothesis testing at a nominal 2α level after each observation has been randomly sampled. Results are given in Table XI.

A Bayesian Approach

For $X \sim \text{Exp}(\lambda)$, $\lambda > 0$ and prior distribution of Λ being $\text{Exp}(\mu)$, μ known (> 0), suppose one is again interested in testing H_0: $\lambda = \lambda_0$ against H_A: $\lambda \neq \lambda_0$ and again without loss of generality one can take $\lambda_0 = 1$ so $2X \sim \chi^2(2)$ under H_0. Measuring the posterior probabilities that λ is less than or greater than 1, a rational and reasonable stopping rule is: sampling stops (with the rejection of H_0) the first time

$$\int_0^1 \pi_1(\lambda).d\lambda < k,$$

or
$$\int_1^\infty \pi_1(\lambda).d\lambda < k, \text{ for some constant } k \in (0,\tfrac{1}{2}), \text{ where } \pi_1$$

is the posterior density of Λ.

TABLE XI

VALUES OF $nx_L^{(n)}$, $nx_U^{(n)}$, (i) $P(\chi^2(2n) < 2nx_L^{(n)})$, (ii) $P(\chi^2(2n) > 2nx_U^{(n)})$, (iii) (i)+(ii) AND P_n FOR $2\alpha = 0.10$

n	$nx_L^{(n)}$	$nx_U^{(n)}$	(i)	(ii)	(iii)	P_n
1	0.083815	3.932144	0.08040	0.01960	0.10	0.10
2	0.414290	5.634473	0.06542	0.02370	0.08912	0.15416
3	0.878496	7.181757	0.05936	0.2583	0.08519	0.18991
4	1.419772	8.646317	0.05598	0.02720	0.08318	0.21649
5	2.012391	10.05722	0.05378	0.02819	0.08197	0.23759
6	2.64230	11.4297				0.25507
7	3.30085	12.7728				0.26996
8	3.98230	14.0927				0.28293
9	4.68257	15.3934				0.29440
10	5.39870	16.6780				0.30468
15	9.15232	22.9268				0.34424
20	13.0989	28.9814				0.37212
30	21.3288	40.7527				0.41084
40	29.8347	52.2475				0.43772
50	38.5176	63.5649				0.45814

Now from Example 4.2,

$$\pi_1(\lambda) = \begin{cases} \frac{(y+\mu)^{n+1}}{n!} \lambda^n e^{-(y+\mu)\lambda}, & y \equiv \sum_{i=1}^{n} x_i, \quad \lambda > 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$\text{so} \int_0^1 \pi_1(\lambda).d\lambda = \frac{1}{n!}\int_0^{y+\mu} u^n e^{-u}.du, \quad u \equiv (y+\mu)\lambda$$

$$= P(\chi^2(2(n+1)) < 2(y+\mu)).$$

$P(\chi^2(2n+2) < 2(y+\mu)$ is a strictly increasing function of $y > 0$ so that the critical region is in the left tail, which agrees with intuition. Moreover, the lower critical point for $\sum_{i=1}^{n} x_i + \mu$ in the nth test is $\frac{1}{2}\chi_k^2(2n+2)$, $n = 1, \ldots$. Similarly,

$$\int_1^{\infty} \pi_1(\lambda).d\lambda = P(\chi^2(2n+2) > 2(\sum_{i=1}^{n} x_i + \mu))$$

is a decreasing function of $\sum_{i=1}^{n} x_i$ (> 0) so that the critical region here is the right tail, which also agrees with intuition. Also, the upper critical point for $\sum_{i=1}^{n} x_i + \mu$ in the nth test is $\frac{1}{2}\chi_{1-k}^2(2n+2)$, $n = 1, \ldots$. (For a vague prior take $\mu = 0$.) Thus the effect this Bayesian approach has on the "original" sampling and testing procedure is to replace the original first sample with μ, suppress the original first test and continue randomly sampling and testing as in the original procedure, the nth actual observation of this Bayesian procedure being included for the first time in the (n+1)th test of the original procecure, $n = 1, \ldots$.

In the case of the vague prior $(\mu = 0)$,

$$P_1 = P(\chi^2(2) \notin (\chi^2_{1-\alpha}(4),\ \chi^2_\alpha(4)))$$

$$= 1 - e^{-y_{11}} + e^{-y_{21}} \quad \text{where} \quad y_{11} = \tfrac{1}{2}\chi^2_{1-\alpha}(4)$$

$$\text{and} \quad y_{21} = \tfrac{1}{2}\chi^2_\alpha\ (4)$$

$$= 0.307785 \quad \text{for} \quad \alpha = 0.05\ .$$

$$P_2 = 1 + e^{-y_{21}} + (y_{11} - y_{12} - 1)e^{-y_{12}} + (y_{21} - y_{11})e^{-y_{22}} \tag{2.11}$$

where y_{11} and y_{12} are as before,

$$y_{12} = \tfrac{1}{2}\chi^2_{1-\alpha}(6)$$

$$\text{and} \quad y_{22} = \tfrac{1}{2}\chi^2_\alpha\ (6)$$

$$= 0.36401 \quad \text{for} \quad \alpha = 0.05\ .$$

Similarly, from (2.18), substituting $y_{13} = \tfrac{1}{2}\chi^2_{1-\alpha}(8)$,

$$y_{23} = \tfrac{1}{2}\chi^2_\alpha\ (8)$$

and $\alpha = 0.05$,

$$P_3 = 0.40785.$$

CHAPTER VI

UNBIASED SEQUENTIAL ESTIMATION

A frequently occurring question which arises naturally after a null hypothesis about a parameter has been rejected is "What then is an estimate of the true value of this parameter?" An often forgotten section of Blackwell's classic publication (11) is that on unbiased sequential estimation. The method will be explained and illustrated.

Suppose $X_1, \ldots, X_n$ are random variables whose distribution depends on parameter θ. If $T(\underset{\sim}{X})$ is unbiased and $U(\underset{\sim}{X})$ sufficient for θ (T with finite variance) then $E(T|U) \equiv V(U)$ is unbiased for θ, depends on only U (not θ) and has variance not greater than that of T with equality if and only if T is a function of U (almost everywhere).

> The estimate obtained in this section for the parameter of a sequential process is of the v type; its importance lies in the fact that in many cases there is an unbiased estimate t (generally poor) which is a function of the first observation, and which will consequently be an unbiased estimate no matter what sequential test procedure is used.

A closed sequential sample (test) is determined by specifying a sequence of mutually exclusive and exhaustive events $\{S_i\}$, where S_i depends on only $x_1, \ldots, x_i$; i.e. $\sum_{i=1}^{\infty} P(S_i) = 1 \; \forall \; \theta$. The event S_i is that sampling stops after the ith observation. Feller (21) has shown that the (test) procedures of Chapters II and III are closed, irrespective of how small α is in the open interval $(0, \frac{1}{2})$. The sequential sampling procedures to follow in illustrating Blackwell's unbiased sequential estimation method are also closed. They are just truncations of the test procedures of Chapters II and III.

Let $\{U_i\}$ denote any sequence of random variables such that $U_i = U_i(X_1, \ldots, X_i)$ is sufficient for estimating θ from $x_1, \ldots, x_i$, and suppose the sequential test (or sample) satisfies the condition $S_i = W_i \cap C(\bigcup_{j=1}^{i-1} S_j)$, where W_i is an event depending on only U_i and $C(A)$ denotes the complement of the event A. This condition means that when the ith observation is taken the decision to stop then depends on only U_i, the value of the ith sufficient statistic. All tests in Chapters II and III satisfy the above condition, as do all sequential sampling procedures to follow in illustrating Blackwell's unbiased sequential estimation method.

Let $\{T_i\}$ denote any sequence of random variables such that $T_i = T_i(X_1, \ldots, X_i)$ and define $T = T_i$ when S_i occurs. Then T is said to be unbiased for θ (relative to the particular sequential test $\{S_i\}$) if and only if $E(T) = \theta \; \forall \; \theta$.

Now let T denote <u>any</u> unbiased estimate of θ relative to a particular sequential test $\{S_i\}$, let h_i denote the indicator function of event $C(\bigcup_{j=1}^{i} S_j)$ and define

$$V = \frac{E(h_{i-1}\, T_i | U_i)}{E(h_{i-1} | U_i)} \quad \text{when } S_i \text{ occurs.}$$

Blackwell (11) has shown V to be unbiased for θ.

There are some important points worth mentioning before proceding to illustrate Blackwell's unbiased sequential estimation method. First is a result due principally to Fay.

<u>Fay's Lemma:</u> If, for each m, $T_m = T_m(X_1, \ldots, X_m)$ is sufficient for θ in the case of the sample $X_1, \ldots, X_m$ of fixed size, then (N, T_N) is sufficient for θ in the sequential case.

Lehmann (35) and Blackwell (11) have given proofs. From Fay's Lemma it follows that if $X_1, \ldots$ are i.i.d. $N(\theta, 1)$ or $Exp(\theta)$ then $(N, \sum_{i=1}^{N} X_i)$ is sufficient for θ.

Second, Lehmann and Stein (36) have shown that the sequential test procedures of Chapters II and III in the normal cases are <u>not</u> complete, i.e. (N, T_N), where $T_N \equiv \sum_{j=1}^{N} X_j$, is <u>not</u> complete in these normal cases. This is also true in more general circumstances involving sequential random sampling from a normal distribution with the trivial exception of (procedures with) fixed sample size. It appears the question of completeness or otherwise of this statistic in the case of sequential random sampling from an underlying exponential distribution is still open.

Now to illustrate Blackwell's unbiased sequential estimation procedure. In both the normal and exponential sequential procedures, $T = T_1 = X_1$ may be taken as an unbiased estimator - for μ in the normal cases and for $\frac{1}{\lambda}$ in the exponential cases - and (N, T_N), where $T_N \equiv \sum_{j=1}^{N} X_j$, may be taken as a statistic sufficient for estimating μ in the normal cases and $\frac{1}{\lambda}$ in the exponential cases from $x_1, \ldots, x_n$ for $n = 1, \ldots$.

Consider the two-tailed normal test procedure with $2\alpha = 0.05$ truncated at $n = 2$. The test procedure is (or was) of the form: Take the first random observation; if it lies outside the interval $(-1.96, 1.96)$ then stop sampling; if it lies <u>in</u> the given interval take a second random observation (independent of the first) and then stop sampling. The "joint density" of (N, X_1, X_2) may be taken as

$$f_n(x_1, x_2) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_1-\mu)^2}, & n = 1,\ |x_1| > 1.96 \text{ and } x_2 = 0 \text{(say!)} \\ & \text{[one-dimensional, i.e. } x_2 \text{ is to be considered degenerate at 0],} \\ \frac{1}{2\pi} e^{-\frac{1}{2}\left[(x_1-\mu)^2 + (x_2-\mu)^2\right]}, & n = 2,\ |x_1| < 1.96 \text{ and } \forall\ x_2 \\ & \text{[two-dimensional],} \\ 0 & \text{otherwise.} \end{cases}$$

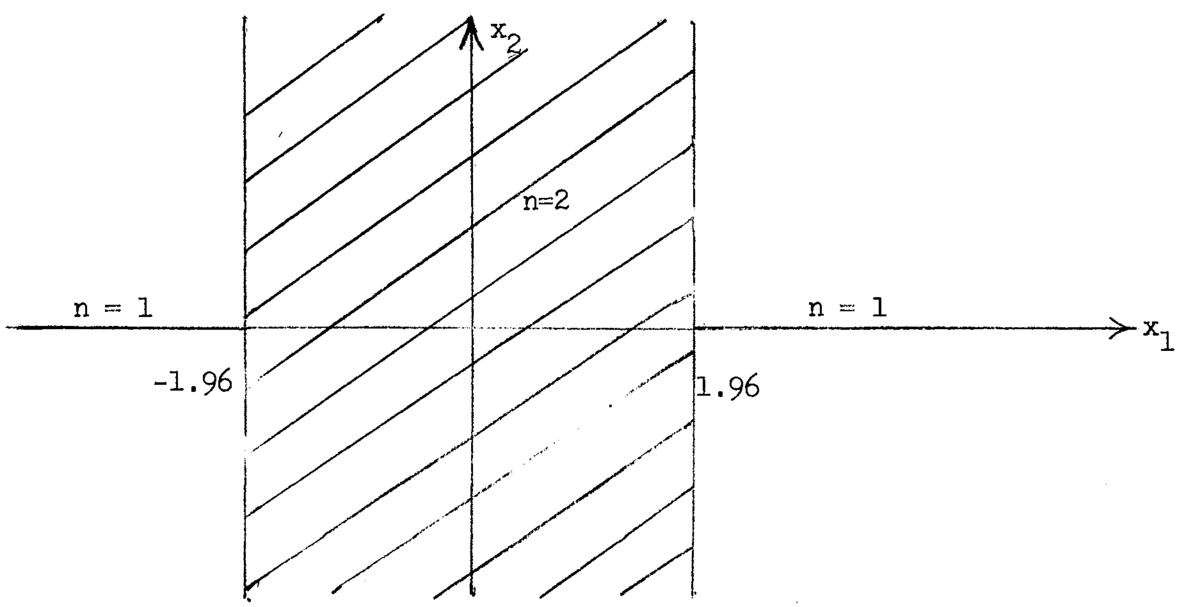

Figure 2. Graph of $f_n(x_1,x_2)$ in Two-tailed Normal Case for $2\alpha = 0.05$ Truncated at $n = 2$

The marginals may then be calculated:

$$P(N=n) = \int_{-\infty}^{+\infty}\int_{\infty}^{\infty} dF_n(x_1,x_2), \quad n = 1, 2$$

$$= \begin{cases} \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{|x_1| > 1.96} e^{-\frac{1}{2}(x_1-\mu)^2}.dx_1, & n = 1, \\ \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-1.96}^{1.96} e^{-\frac{1}{2}(x_1-\mu)^2}.dx_1, & n = 2 \end{cases}$$

$$f_{X_1}(x_1) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_1-\mu)^2}$$

$$f_{X_2}(x_2) \equiv \left[\frac{1}{\sqrt{2\pi}}\int_{-1.96}^{1.96} e^{-\frac{1}{2}(x_1-\mu)^2}.dx_1\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_2-\mu)^2}$$

Note that $\displaystyle\int_{-\infty}^{\infty} f_{X_2}(x_2).dx_2 = P(|X_1| < 1.96)$

$$= P(N = 2)$$

$$\lneq 1.$$

For $n = 2$, $v \equiv \dfrac{E(h_1(X_1)X_1 | X_1+X_2 = u_2)}{E(h_1(X_1) | X_1+X_2 = u_2)}$

$$\equiv \frac{\int_{-1.96}^{1.96} x_1 f_2(x_1, x_2 | X_1 + X_2 = u_2).dx_1}{\int_{-1.96}^{1.96} f_2(x_1, x_2 | X_1 + X_2 = u_2).dx_1} ,$$

$f_2(x_1, x_2 | X_1 + X_2 = u_2)$

$$\equiv \begin{cases} \dfrac{f_2(x_1,\ u_2 - x_1)}{g_{x_1+x_2}(u_2)} , & |x_1| < 1.96, \\ 0 & \text{otherwise,} \end{cases}$$

where $f_2(x_1,\ u_2 - x_1) = \dfrac{1}{2\pi} e^{-\frac{1}{2}\left[(x_1 - \mu)^2 + (u_2 - x_1 - \mu)^2\right]}, \ |x_1| < 1.96$

$$= \begin{cases} \dfrac{1}{2\pi} e^{-\frac{1}{2}(2x_1^2 + 2\mu^2 + u_2^2 - 2u_2x_1 - 2u_2\mu)}, \\ \hfill |x_1| < 1.96, \\ 0 \quad \text{otherwise,} \end{cases}$$

and $g_{x_1+x_2}(u_2) \equiv \int_{-\infty}^{\infty} f_2(x_1,\ u_2 - x_1).dx_1$

$$\equiv \frac{1}{2\pi} e^{-\frac{1}{2}(u_2^2 - 2u_2\mu + 2\mu^2)} \int_{-1.96}^{1.96} e^{-(x_1^2 - u_2x_1)}.dx_1 \ .$$

Now $\int_{-1.96}^{1.96} e^{-(x_1^2 - u_2x_1)}.dx_1 = e^{\frac{1}{4}u_2^2} \int_{-1.96}^{1.96} e^{-(x_1 - \frac{1}{2}u_2)^2}.dx_1$

$$= e^{\frac{1}{4}u_2^2} \frac{1}{\sqrt{2}} \int_{-\sqrt{2}(1.96 + \frac{1}{2}u_2)}^{\sqrt{2}(1.96 - \frac{1}{2}u_2)} e^{-\frac{1}{2}w^2}.dw, \quad w \equiv \sqrt{2}(x_1 - \tfrac{1}{2}u_2)$$

$$= \sqrt{\pi}\, e^{\frac{1}{4}u_2^2}\left[\Phi(\sqrt{2}(1.96 - \tfrac{1}{2}u_2)) - \Phi(-\sqrt{2}(1.96 + \tfrac{1}{2}u_2))\right]$$

so $$g_{x_1+x_2}(u_2) \equiv \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}u_2^2 + u_2\mu - \mu^2} \left[\ \cdot\ \right];$$

$$f_2(x_1,x_2|X_1+X_2 = u_2) = \frac{1}{\sqrt{\pi}} \left[\ \cdot\ \right]^{-1} e^{-(x_1-\frac{1}{2}u_2)^2}, \quad |x_1| < 1.96,$$

and $$v = \frac{\int_{-1.96}^{1.96} x_1 e^{-(x_1-\frac{1}{2}u_2)^2}.dx_1}{\int_{-1.96}^{1.96} e^{-(x_1-\frac{1}{2}u_2)^2}.dx_1}$$

$$= \frac{\int_{-1.96}^{1.96} (x_1-\frac{1}{2}u_2)e^{-(x_1-\frac{1}{2}u_2)^2}.dx_1 + \frac{1}{2}u_2\int_{-1.96}^{1.96} e^{-(x_1-\frac{1}{2}u_2)^2}.dx_1}{\sqrt{\pi}\ \left[\ \cdot\ \right]}$$

$$= \frac{1}{2}u_2 + \frac{e^{-(1.96+\frac{1}{2}u_2)^2} - e^{-(1.96-\frac{1}{2}u_2)^2}}{2\sqrt{\pi}\ \left[\Phi(\sqrt{2}(1.96-\frac{1}{2}u_2)) - \Phi(-\sqrt{2}(1.96+\frac{1}{2}u_2))\right]} .$$

Clearly this illustration may be generalized to values of 2α other than 0.05 and to one-tailed test procedures truncated at $n = 2$.

Consider now an exponential test procedure truncated at $n = 2$. The test procedure is (or was) of the form: Take the first random observation; if it lies outside the interval (a,b) then stop sampling; if it lies in the given interval take a second random observation (independent of the first) and then stop sampling. Critical points a and b are subject to only $0 \leq a < b$. For a right-tailed test $a = 0$. For a left-tailed test take $b = \infty$. The "joint density" of (N, X_1, X_2) may be taken as

$$
f_n(x_1,x_2) = \begin{cases} \lambda e^{-\lambda x_1}, \quad n = 1, \quad x_1 \notin (a,b),\ x_1 > 0 \text{ and } x_2 = 0 \text{ (say!)} \\ \quad [\text{one-dimensional, i.e. } x_2 \text{ is to be considered degenerate at } 0], \\ \lambda^2 e^{-\lambda(x_1+x_2)}, \quad n = 2, \quad x_1 \in (a,b) \text{ and } x_2 > 0 \\ \quad [\text{two-dimensional}], \\ 0 \quad \text{otherwise.} \end{cases}
$$

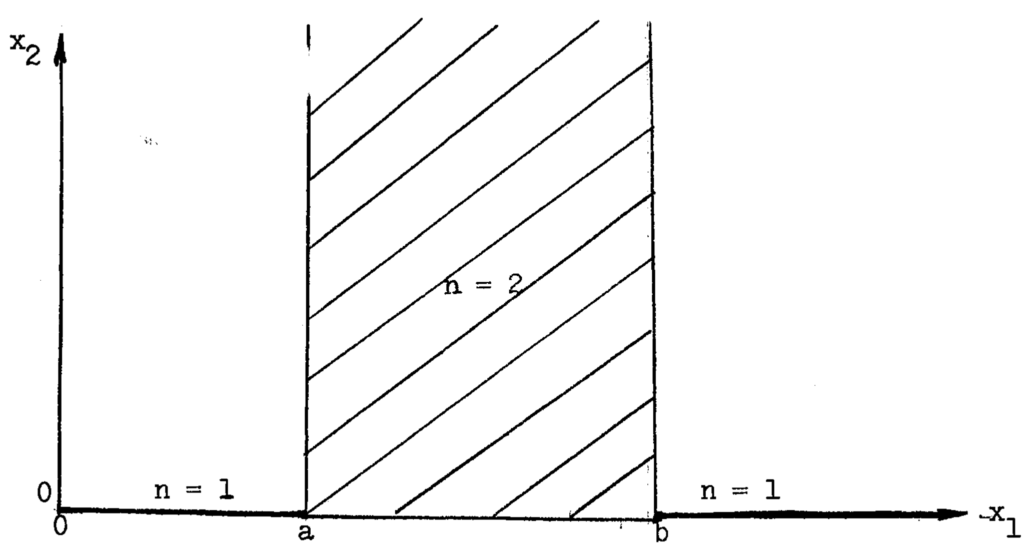

Figure 3. Graph of $f_n(x_1,x_2)$ in Exponential Cases Truncated at $n = 2$

The marginals then follow:

$$P(N=n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dF_n(x_1,x_2) \quad , \quad n = 1, 2$$

$$= \begin{cases} \lambda \displaystyle\int_{\substack{x_1>0, \\ x_1 \notin (a,b)}} e^{-\lambda x_1}.dx_1, & n = 1, \\ \\ \lambda \displaystyle\int_a^b e^{-\lambda x_1}.dx_1, & n = 2, \end{cases}$$

$$= \begin{cases} 1 - (e^{-\lambda a} - e^{-\lambda b}) & , \quad n = 1, \\ e^{-\lambda a} - e^{-\lambda b} & , \quad n = 2. \end{cases}$$

$$f_{x_1}(x_1) = \begin{cases} \lambda\, e^{-\lambda x_1} & , \quad x_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_{x_2}(x_2) = \left[\lambda \int_a^b e^{-\lambda x_1}.dx_1\right] e^{-\lambda x_2} \quad , \quad x_2 > 0,$$

$$= \begin{cases} (e^{-\lambda a} - e^{-\lambda b})\, e^{-\lambda x_2} & , \quad x_2 > 0, \\ 0 \quad \text{otherwise.} \end{cases}$$

Note that $\displaystyle\int_{-\infty}^{\infty} f_{x_2}(x_2).dx_2 = P(N=2).$

For $n = 2$, $v = \dfrac{E(h_1(X_1)X_1 \mid X_1+X_2 = u_2)}{E(h_1(X_1) \mid X_1+X_2 = u_2)}$ is unbiased for $(\frac{1}{\lambda})$.

$$= \left\{ \frac{\int_a^b x_1 f_2(x_1,x_2|X_1+X_2 = u_2).dx_1}{\int_a^b f_2(x_1,x_2|X_1+X_2 = u_2).dx_1} \right. ,$$

$$f_2(x_1,x_2\ X_1+X_2 = u_2) = \begin{cases} \dfrac{f_2(x_1,u_2-x_1)}{g_{x_1+x_2}(u_2)} , & x_1 \varepsilon\ (a,b),\ x_1 < u_2 , \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{where}\quad f_2(x_1,u_2-x_1) = \begin{cases} \lambda^2\ e^{-\lambda u_2}, & x_1 \varepsilon\ (a,b),\ x_1 < u_2, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and}\quad g_{x_1+x_2}(u_2) \equiv \int_{-\infty}^{\infty} f_2(x_1,u_2-x_1).dx_1$$

$$= \begin{cases} 0, & u_1 < a, \\ \lambda^2(u_2-a)e^{-\lambda u_2}, & u_2\ \varepsilon\ (a,b), \\ \lambda^2(b-a)e^{-\lambda u_2}\ , & u_2 > b; \end{cases}$$

$$\therefore f_2(x_1,x_2|X_1+X_2 = u_2) = \begin{cases} \dfrac{1}{u_2-a}\ , & x_1\ \varepsilon\ (a,u_2),\ u_2\ \varepsilon\ (a,b), \\ \dfrac{1}{b-a}\ , & x_1\ \varepsilon\ (a,b),\ u_2 > b, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{and}\quad v = \begin{cases} \dfrac{\frac{1}{u_2-a}\int_a^{u_2} x_1.dx_1}{\frac{1}{u_2-a}\int_a^{u_2} dx_1}\ , & u_2 < b, \\ \dfrac{\frac{1}{b-a}\int_a^{b} x_1.dx_1}{\frac{1}{b-a}\int_a^{b} dx_1}\ , & u_2 > b \end{cases}$$

$$= \begin{cases} \frac{1}{2}(a+u_2), & u_2 < b, \\ \frac{1}{2}(a+b), & u_2 > b. \end{cases}$$

Verification of unbiasedness:

$$E(V) = \lambda\int_0^a x_1 e^{-\lambda x_1}.dx_1 + \lambda\int_b^\infty x_1 e^{-\lambda x_1}.dx_1$$

$$+ \tfrac{1}{2}\lambda^2\int_a^b (u_2^2-a^2)e^{-\lambda u_2}.du_2 + \tfrac{1}{2}\lambda^2(b^2-a^2)\int_b^\infty e^{-\lambda u_2}.du_2$$

$$= \frac{1}{\lambda} \quad .$$

Variance of V:

$$E(V^2) = \lambda\int_0^a x_1^2 \, e^{-\lambda x_1}.dx_1 + \lambda\int_b^\infty x_1^2 \, e^{-\lambda x_1}.dx_1$$

$$+ \tfrac{1}{4}\lambda^2\int_a^b (u_2^3 + au_2^2 - a^2u_2 - a^3) \, e^{-\lambda u_2}.du_2$$

$$+ \tfrac{1}{4}\lambda^2(b^3 + ab^2 - a^2b - a^3)\int_b^\infty e^{-\lambda u_2}.du_2$$

$$= \frac{1}{4\lambda^2}\,(8 - 2e^{-\lambda a} + (\lambda^2n^2+2\lambda b-2\lambda^2ab+2-2\lambda a + \lambda^2a^2)e^{-\lambda b})$$

so $\quad var(V) = \dfrac{1}{4\lambda^2}\,(4-2e^{-\lambda a} + ([\lambda(b-a) + 1]^2 + 1)e^{-\lambda b}).$

If the test procedures are extended to taking a third sequential observation then Blackwell's method above becomes very complex and "untidy".

CHAPTER VII

AN OVERVIEW, SUMMARY AND EXTENSIONS

Surely the prime motivation for Wald and others to develop the Sequential Probability Ratio Test (SPRT) was to provide a sequential analysis of data as it is accumulated with a test which has prechosen overall probabiliteis of Types I and II errors, or at least excellent approximations thereto. This test may be used to advantage in cases where it is "costly" to take a random sample of prefixed size - particularly when there is no guarantee that this fixed-sample-size procedure will yield conclusive results, or the action to be taken is dictated in a fraction of the prefixed sample size and sampling is continued only to vainly satisfy the conditions and properties of the preconceived sampling scheme. It is the economics (or tedium) of a context that most often forces an experimenter to use a sequential acheme.

This dissertation has tackled a slightly different problem. An experimenter may be interested in "legitimately" discounting a certain (null) hypothetical claim and to do so runs an experiment, which yields what is considered "insufficient statistical evidence" (in the form of an observation on a test statistic) against the claim. The experimenter repeats the experiment enough times to collect "sufficient statistical evidence" to refute the claim. Qualitatively, the probability of Type I error rises above the nominal value at which successive combinations of observations on the test statistic may have been tested.

(Often an experimenter in an applied field, using statistics as only a tool, is not consciously aware of this fundamental qualitative result. In view of the experimenter's unwillingness to change his system, the next best approach a theoretical statistician can adopt is to determine just what it is that the experimenter is <u>really</u> doing - what are the true frequency characteristics of the sequential scheme the experimenter is following.) This dissertation has gone some of the way towards answering how this rise takes place quantitatively: "the answer" depends on both the distributional form of the underlying test statistic and the mode of combination.

The only underlying test statistics considered in this dissertation are the only two continuous statistics that Armitage and McPherson considered: normal and exponential (equivalent to a chi-squared with two degrees of freedom). The computational advantages are immediate: linear combinations of normal variates are normal and sums of independent exponentials are within a constant multiple of chi-squared distributions with an even number of degrees of freedom. Moreover, if the underdlying test statistic is not one of these two distributional forms, then it may be converted to a chi-squared variate with two degrees of freedom by "Fisher's transformation" $(\cdot \to -2\ln \cdot)$ applied to the significance level of the original statistic, assuming the original statistic is continuous. (If the original statistic is discrete then modified methods - Lancaster's approximation (31, 33) in particular - may be employed.)

In its original form Wald's SPRT has an immediate major drawback: while it is certain that the test will terminate (with a finite sample size) there is no upper limit on the sample size required for termination. Understandably, manufacturers (for example) may not be prepared to permit unlimited sampling from their wares, particularly in view of the fact that the cases where "large" sample sizes are likely to be encountered are when the (simple) hypothetical claims being weighed against each other are "close together" - where, due to variation, sample differences tend to be non-significant and population differences tend to be insignificant from a practical viewpoint. Thus a form of truncation is desirable and, as referenced in the problem stated at the end of Chapter I, some research has been done on some truncated SPRT's in exponential testing. Potential truncation possibilities for the general sequential method employed in this dissertation are evident from Chapters II and III for pre-specified simple "null" and alternative hypotheses and for prespecified overall probabilities of Types I and II errors (as in Wald's SPRT) about a normal mean with known variance (i.e. no nuisance parameter) and exponential parameter - in the form of a maximum number (n_{Max}) of observations to be randomly sampled (40, 41).

Wald and Wolfowitz (50) have shown that the SPRT has an optimal property: "of all tests with the same power the sequential probability ratio test requires on the average fewest observations." In contrast, Gundy and Siegmund (25) have shown that if $X_1, \ldots, X_n \sim$ i.i.d. (0, 1), i.e. zero mean and unit variance, $S_n \equiv \sum_{i=1}^{n} X_i$, $n = 1, \ldots,$ and t_c denotes the smallest integer n such that $|S_n| > c\, n^{\frac{1}{2}}$ ($= \infty$ if no such n exists), $c \geq 0$, then $E(T_c) < \infty$ if $0 \leq c < 1$; $E(T_c) = \infty$ if $c \geq 1$. (Clearly the result can be generalized to any i.i.d. variates X_i which possess a non-zero and finite variance.) Thus no sampling scheme considered in Chapter II has a finite average sample number.

It may be of interest to compare the tables generated by the two underlying distributions considered here with tables generated by other distributions underlying the general fixed-sample-size procedure adopted here. Distributions of immediate interest include chi-squared distributions (more generally gammas), the Laplace (double exponential) distribution, Weibull distributions, Student's T (40), Snedecor's F and multivariate distributions.

A SELECTED BIBLIOGRAPHY

(1) Anscombe, F. J. "Dependence of the Fiducial Argument on the Sampling Rule." Biometrika, Vol. 44 (1957), 464-469.

(2) Anscombe, F. J. "Fixed-Sample-Size Analysis of Sequential Observations." Biometrika, Vol. 10 (1954), 89-100.

(3) Armitage, P., C. K. McPherson, and B. C. Rowe. "Repeated Significance Tests on Accumulating Data." Journal of the Royal Statistical Society, Series A, Vol. 132 (1969), 235-244.

(4) Aroian, L. A. "Exact Truncated Sequential Tests for the Exponential Density Function." Proceedings of the Ninth National Symposium on Reliability and Quality Control, 1963, 470-486.

(5) Aroian, L. A. "Sequential Analysis, Direct Method." Technometrics, Vol. 10 (1968), 125-132.

(6) Barnard, G. A. Discussion on Kalbfleisch and Sprott's paper, "Applications of Likelihood Methods to Models involving Large Numbers of Parameters." Journal of the Royal Statistical Society, Series B, Vol. 32 (1970), 175-208.

(7) Barnard, G. A. "The Meaning of a Significance Level." Biometrika, Vol. 34 (1947), 179-182.

(8) Barnard, G. A. Review of "Sequential Analysis" by Abraham Wald. Journal of the American Statistical Association, Vol. 42 (1947), 658-669.

(9) Barraclough, E. D. and E. S. Page. "Tables for Walt Tests for the Mean of a Normal Distribution." Biometrika, Vol. 46 (1959), 169-177.

(10) Bhate, D. H. "A Note on the Distribution of Successive Sums of Samples from an Exponential Population." Bulletin of the Calcutta Statistical Association, Vol. 8 (1958), 13-19.

(11) Blackwell, D. "Conditional Expectation and Unbiased Sequential Estimation." Annals of Mathematical Statistics, Vol. 18 (1947), 105-110.

(12) Box. G. E. P. and G. C. Tiao. Bayesian Inference in Statistical Analysis. Reading, Massachusetts: Addison-Wesley, 1973.

(13) Burman, J. P. "Sequential Sampling Formulae for a Binomial Population." Journal of the Royal Statistical Society, Supplement, Vol. 8 (1946), 98-103.

(14) Burnett, T. L. "Truncation of Sequential Life Tests." Proceedings of the Eighth National Symposium on Reliability and Quality Control, 1962, 7-13.

(15) Cornfield, J. "The Bayesian Outlook and its Application" (with Discussion). Biometrics, Vol. 25 (1969), 617-657.

(16) Cornfield, J. "Sequential Trials, Sequential Analysis and the Likelihood Principle." American Statistician, Vol. 20 (1966), 18-23.

(17) Cox, D. R. "Some Problems Connected with Statistical Inference." Annals of Mathematical Statistics, Vol 29 (1958), 357-372.

(18) Easterling, R. G. "A Personal View of the Bayesian Controversy in Reliability and Statistics." Institute of Electrical and Electronics Engineers Transactions on Reliability, Vol. R-21 (1972), 186-194.

(19) Epstein, B. "Truncated Life Tests in the Exponential Case." Annals of Mathematical Statistics, Vol. 25 (1954), 555-564.

(20) Epstein, B. and M. Sobel. "Sequential Life Tests in the Exponential Case." Annals of Mathematical Statistics, Vol. 26 (1955), 82-93.

(21) Feller, W. "The General Form of the S-called Law of the Iterated Logarithm." Transactions of the American Mathematical Society, Vol. 54 (1943), 373-402.

(22) Fisher, R. A. Statistical Methods for Research Workers, 10th ed. London: Oliver and Boyd, 1946.

(23) Fisz, M. Probability Theory and Mathematical Statistics, 3rd ed. New York: Wiley, 1963.

(24) Good, I. J. "On the Weighted Combination of Significance Tests." Journal of the Royal Statistical Society, Series B, Vol. 17 (1955), 264-265.

(25) Gundy, R. E. and D. Siegmund. "On a Stopping Rule and the Central Limit Theorem." Annals of Mathematical Statistics, Vol. 38 (1967), 1915-1917.

(26) Hacking, I. Logic of Statistical Inference. Cambridge University Press, 1965.

(27) Jeffreys, H. Theory of Probability, 3rd ed. Oxford: Clarendon Press, 1961.

(28) Kalbfleisch, J. D. and D. A. Sprott. "Applications of Likelihood Methods to Models involving Large Numbers of Parameters" (with Discussion). Journal of the Royal Statistical Society, Series B, Vol. 32 (1970), 175-208.

(29) Kempthorne, O. "Theories of Inference and Data Analysis." In Statistical Papers in Honor of George W. Snedecor. Ames, Iowa: Iowa State University Press, 1972.

(30) Kempthorne, O. and J. L. Folks. Probability, Statistics, and Data Analysis. Ames, Iowa: Iowa State University Press, 1971.

(31) Kincaid, W. M. "The Combination of Tests Based on Discrete Distributions." Journal of the American Statistical Association, Vol. 57 (1962), 10-19.

(32) Lancaster, H. O. "The Combination of Probabilities: An Application of Orthonormal Functions." Australian Journal of Statistics, Vol. 3 (1961), 20-33.

(33) Lancanster, H. O. "The Combination of Probabilities Arising from Data in Discrete Distributions." Biometrika, Vol. 36 (1949), 370-382.

(34) Lehmann, E. L. Testing Statistical Hypotheses. New York: Wiley, 1959.

(35) Lehmann, E. L. Theory of Estimation and Testing Hypotheses. (Mimeographed notes, Associated Students Store, University of California, Berkeley.)

(36) Lehmann, E. L. and C. Stein. "Completeness in the Sequential Case." Annals of Mathematical Statistics, Vol. 21 (1950), 376-385.

(37) Lindley, D. V. Discussion on Dr. Copas's paper, "Compound Decisions and Empirical Bayes." Journal of the Royal Statistical Society, Series B, Vol. 31 (1969), 419-421.

(38) Littell, R. C. and J. L. Folks. "Asymptotic Optimality of Fisher's Method of Combining Independent Tests." Journal of the American Statistical Association, Vol. 66 (1971), 802-806.

(39) Littell, R. C. and J. L. Folks. "Asymptotic Optimality of Fisher's Method of Combining Independent Tests II." Journal of the American Statistical Association, Vol. 68 (1973), 193-194.

(40) McPherson, C. K. Some Problems in Sequential Experimentation. (Unpublished PhD thesis, University of London, 1972.)

(41) McPherson, C. K. and P. Armitage. "Repeated Significance Tests on Accumulating Data when the Null Hypothesis is not True." Journal of the Royal Statistical Scoiety, Series A, Vol. 134 (1971), 15-25.

(42) Merrington, M. "Numerical Approximation to the Percentage Points of the χ^2 Distribution." Biometrika, Vol. 32 (1941), 200-202.

(43) Novick, M. R. "Multiparameter Bayesian Indifference Procedures" (with Discussion). Journal of the Royal Statistical Society, Series B, Vol. 31 (1969), 29-64.

(44) Pearson, E. S. "On Questions Raised by the Combination of Tests Based on Discontinuous Distributions." Biometrika, Vol. 37 (1950), 383-398.

(45) Savage, L. J. "The Foundations of Statistics Reconsidered," 1961, In Studies in Subjective Probability, edited by Henry E. Kyburg, Jr. and Howard E. Smokler. New York: Wiley, 1964, 173-188.

(46) Stein, C. "A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance." Annals of Mathematical Statistics, Vol. 16 (1945), 243-258.

(47) Thompson, C. M. "Tables of the Percentage Points of the χ^2-Distribution." Biometrika, Vol. 32 (1941), 188-191.

(48) VanZwet, W. R. and J. Oosterhoff. "On the Combination of Independent Test Statistics." Annals of Mathematical Statistics, Vol. 38 (1967), 659.

(49) Wald, A. Sequential Analysis. New York: Wiley, 1947.

(50) Wald, A. and J. Wolfowitz. "Optimum Character of the Sequential Probability Ratio Test." Annals of Mathematical Statistics, Vol. 19 (1948), 326-339.

(51) Wallis, W. A. "Compounding Probabilities from Independent Significance Tests." Econometrika, Vol. 10 (1942), 229-249.

(52) Wilson, E. B. and M. M. Hilferty. Proceedings of the National Academy of Science, Washington, Vol. 17 (1931), 684.

(53) Woodall, R. C. and B. M. Kurkjian. "Exact Operating Characteristic for Truncated Sequential Life Tests in the Exponential Case." Annals of Mathematical Statistics, Vol. 33 (1962), 1403-1412.

(54) Zelen, M. and L. S. Joel. "The Weighted Compounding of Two Independent Significance Tests." Annals of Mathematical Statistics, Vol. 30 (1959), 885-895.

VITA

Malcolm Ross Heyworth

Candidate for the Degree of

Doctor of Philosophy

Thesis: SEQUENTIAL SIGNIFICANCE TESTING AND ESTIMATION

Major Field: Statistics

Biographical:

Personal Data: Born in Hamilton, New Zealand, July 21, 1946, youngest of the three sons of Sydney Bleakley and Catherine Hilda Heyworth.

Education: Attended primary school in Hamilton, New Zealand, received secondary schooling at Hamilton Technical College (now Fraser High School), 1960-1963 (Dux, 1963), and Hamilton Boys' High School, 1964 (Proxime Accessit and National Scholar); graduated Bachelor of Science in pure and applied mathematics from the University of Auckland, Auckland, New Zealand, May, 1969, and Master of Science (Honours) in mathematics from the same institution, May, 1970; requirements for the Degree of Doctor of Philosophy completed at Oklahoma State University, Stillwater, Oklahoma, July, 1974.

Professional Experience: Tutor in the Department of Mathematics, University of Auckland, New Zealand, 1966 and 1968; Assistant in the Department of Mathematics, Westlake Boys' High School, Westlake, Auckland, February through August, 1970; Graduate Teaching Assistant in the Department of Mathematics and Statistics (Department of Statistics after July 1, 1973), Oklahoma State University, Stillwater, Oklahoma, August, 1970 to May, 1974.