4

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0361329 4

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

Copy 4

ORNL-2629
Physics and Mathematics

SOME PROBLEMS IN STATISTICAL INFERENCE

J. J. Gart

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this document, send in name with document and the library will arrange a loan.

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

Printed in USA. Price 275 cents. Available from the

Office of Technical Services
U. S. Department of Commerce
Washington 25, D. C.

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission to the extent that such employee or contractor prepares, handles or distributes, or provides access to, any information pursuant to his employment or contract with the Commission.

ORNL-2629

Contract No. W-7405-eng-26

MATHEMATICS PANEL

SOME PROBLEMS IN STATISTICAL INFERENCE

J. J. Gart*

ORINS - AEC Fellowship. Now located in the Department of Biostatistics, Johns Hopkins University, Baltimore, Md.

DATE ISSUED

DEC 10 1958

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0361329 4

TABLE OF CONTENTS

PAGE

I. AN EXTENSION OF THE CRAMER-RAO INEQUALITY 10

1.1 Review of the Literature and Summary 10

1.2 Notation 12

1.3 The Continuous Case. 12

1.4 The Discrete Case. 15

1.5 Conditions for Equality. 16

1.6 Most Unfavorable Distributions 22

1.7 Most Unfavorable Distributions for Some Laplacian Distributions. 23

1.8 More Stringent Inequalities. 27

1.9 The Sequential Case. 30

1.10 Linear Estimation of $E(\theta_k)$ 33

II. THE ASYMPTOTIC PROPERTIES OF ML ESTIMATES WHEN SAMPLING FROM ASSOCIATED POPULATIONS 36

2.1 Summary. 36

2.2 Notation and Assumptions 37

2.3 Consistency. 39

2.4 Asymptotic Normality 44

2.5 Asymptotic Distribution of the Likelihood Ratio. . 46

2.6 An Application 49

III. A SEQUENTIAL DECISION PROCEDURE FOR COMPARING SURVIVAL CURVES. 51

3.1 Introduction 51

PAGE

III. (Continued)

3.2 Review of Literature 53

3.3 Distribution Theory. 54

3.4 Requirements of the Survival Curve 58

3.5 General Statement of the Decision Procedure. . . . 59

3.6 The $f(a_r; \lambda, s)$ when $p(t) = e^{k(t-b)^2}$ 61

3.7 Discussion of the Logistic Curve 66

3.8 Derivation of $f(a_r; u, s)$ for the Logistic Model . 68

3.9 A Linear Approximation to the Survival Curve . . . 76

3.10 Derivation of $f(a_r; p, s)$ for the Linear Model . . 79

3.11 Robustness of the Procedure with Respect to the Form of the Survival Curves. 84

3.12 Comparison of the Original Analysis to the Logistic 91

3.13 The OC Function. 93

3.14 The ASN Function 97

3.15 The Effect of Truncation 100

3.16 An Illustration of the Procedure 103

3.17 Possible Extensions. 105

3.18 Summary. 105

IV. ACKNOWLEDGMENTS. 106

V. BIBLIOGRAPHY 107

VI. APPENDICES 111

VI. (Continued) PAGE

Appendix A. The Covariance in Terms of Conditional Expectations 111

Appendix B. A Complex Integration Theorem. 111

Appendix C. Explicit Formulas for $f_L(a_r; u, s)$ for $a_r = 0, 1, \ldots, 2s+1$ and $s = 0, 1, 2$ . . . 117

LIST OF TABLES

PAGE

Table 1. Table of $f_L(a_r; p, s)$ for $s = 0, 1, 2$ 77

Table 2. Table of $f_U(a_r; p, s)$ for $s = 0, 1, 2$ 83

Table 3. Table of $\alpha^*(\alpha, \beta)$ and $\beta^*(\alpha, \beta)$ for $p_0 = 0.6$, $p_1 = 0.4$; and $s = 0, 1, 2$ 88

Table 4. Table of $\alpha^*(\alpha, \beta)$ and $\beta^*(\alpha, \beta)$ for $p_0 = 0.7$, $p_1 = 0.3$; and $s = 0, 1, 2$ 89

Table 5. Table of the OC and ASN Functions for $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; and $s = 0, 1, 2$ 95

Table 6. Upper bounds on $\alpha(n_0)$ and $\beta(n_0)$ for $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; and $s = 0, 1, 2$ 102

Table 7. The Decision Procedure for Cysteine Hydrochloride . . 104

LIST OF FIGURES

PAGE

Figure 1. The OC Function for $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; and $s = 0, 1, 2$ 96

Figure 2. The ASN Function for $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; and $s = 0, 1, 2$ 99

Figure 3. The Contour of Integration. 113

I. AN EXTENSION OF THE CRAMER-RAO INEQUALITY

1.1 Review of the Literature and Summary. Cramer [6], p. 474 ff., and Rao [12] derived independently a lower bound for the mean square error of an estimate t of a parameter α which appears in a frequency function of a specified form. This expression, alternately termed the Cramer-Rao inequality or the information limit, is

$$(1.1.1) \qquad E(t-\alpha)^2 \geq [E(t) - \alpha]^2 + \frac{\left[\frac{\partial E(t)}{\partial \alpha}\right]^2}{E\left(\frac{\partial \ln \phi}{\partial \alpha}\right)^2},$$

where ϕ is the likelihood of the sample. The expression $E(\partial \ln\phi/\partial\alpha)^2$ is called the information on α and is sometimes denoted by $I(\alpha)$. Under rather general conditions it can be shown equal to $E(-\partial^2 \ln\phi/\partial\alpha^2)$.

The equality in (1.1.1) is reached if and only if

$$(1.1.2) \qquad \phi = \phi_1 \quad e^{t\,V(\alpha) + W(\alpha)}$$

where t and ϕ_1 are functions of the observations alone and $V(\alpha)$ and $W(\alpha)$ are functions of α alone. By the results of Pitman [11] and Koopman [10], the form of (1.1.2) implies that t must be a sufficient statistic. The fact that this form of the likelihood yields a minimum variance estimate was first pointed out by Aitken and Silverstone [1]. If we have n observations which are independently and identically distributed, the frequency function of the underlying population must be of the so-called Pitman-Koopman form,

$$f(x;\alpha) = u(\alpha)\ h(x)\ e^{P(\alpha)\ g(x)} \tag{1.1.3}$$

and t must be a function of $\sum_{i=1}^{n} g(x_i)$ for the equality in (1.1.2) to hold.

Several extensions of the basic inequality have been derived. Bhattacharyya [4] and Chapman and Robbins [5] have derived results which yield more stringent inequalities in certain instances. Wolfowitz [18] has extended the result to sequential sampling situations. Cramer [7] and Barankin [2] have considered joint bounds on sets of estimates of parameters and Hammersley [9] has derived a lower bound of the mean square error of an estimate for the situation in which the parameter to be estimated can only assume discrete values. Barankin [3] has also considered lower bounds on the general absolute central moments of the estimate.

All these results assume that the parameters involved are constants. Here we shall consider the case where the parameters are random variables. Thus the lower bound of the mean square error of an estimate will take into account the variability due to both the observations and the parameters involved. Necessary and sufficient conditions for equality of the extended inequality are derived. Most unfavorable distributions, i.e., distributions which maximize the lower bound, are defined, and several examples are given. Extensions analogous to those of Bhattacharyya [4] and Wolfowitz [18] are also considered. Finally, bounds on the variance of linear estimates of the mean of the parameter are derived.

1.2 Notation. Consider a frequency function $f(x|\Theta)$, where $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_s)$, the function being specified when Θ is specified. Further, Θ is a random variable having the distribution $G(\Theta)$ defined over a non-degenerate range A_s. Let $X = (x_1, x_2, \ldots, x_n)$ be a random sample from a randomly chosen population having the specified frequency function. Let $t_k = t_k(X)$ be an estimate of Θ_k, $1 \leq k \leq s$, functionally independent of Θ_k. Denote $E(t_k|\Theta)$ by $\gamma_k(\Theta)$ and the conditional likelihood of the sample by $\phi(X|\Theta)$, which in general will be $\prod_{i=1}^{n} f(x_i|\Theta)$.

1.3 The Continuous Case. If $f(x|\Theta)$ is a density, assume $\partial\phi/\partial\Theta_k$ exists for all Θ in A_s and $|\partial\phi/\partial\Theta_k| < H(X)$ where H and $t_k H$ are integrable over R_n, the range of X, which is independent of Θ_k. We have

$$1 = \int_{R_n} \phi dX \tag{1.3.1}$$

and

$$\gamma_k(\Theta) = \int_{R_n} t_k \phi dX \ . \tag{1.3.2}$$

By the assumptions just made (see Cramer [6], p. 66 and p. 475), we may differentiate under the integral signs in (1.3.1) and (1.3.2) and obtain

$$0 = \int_{R_n} \frac{\partial\phi}{\partial\Theta_k} dX = \int_{R_n} \frac{\partial \ln\phi}{\partial\Theta_k} \phi dX \tag{1.3.3}$$

and

$$(1.3.4) \qquad \frac{\partial \psi_k(\theta)}{\partial \theta_k} = \int_{R_n} t_k \frac{\partial \phi}{\partial \theta_k} dX = \int_{R_n} t_k \frac{\partial \ln \phi}{\partial \theta_k} \phi dX .$$

Finding expectations of (1.3.3) and (1.3.4) with respect to θ, we have

$$(1.3.5) \qquad 0 = \int_{A_s} \int_{R_n} \frac{\partial \ln \phi}{\partial \theta_k} \phi dX \, dG(\theta)$$

and

$$(1.3.6) \qquad E\left(\frac{\partial \psi_k(\theta)}{\partial \theta_k}\right) = \int_{A_s} \frac{\partial \psi_k(\theta)}{\partial \theta_k} dG(\theta) = \int_{A_s} \int_{R_n} t_k \frac{\partial \ln \phi}{\partial \theta_k} \phi dX \, dG(\theta).$$

By the Schwarz inequality we may write

$$(1.3.7) \left\{ \int_{A_s} \int_{R_n} \left[t_k - \theta_k - E\left[\psi_k(\theta)\right] + E(\theta_k)\right]^2 \phi dX \, dG(\theta) \right\} \left\{ \int_{A_s} \int_{R_n} \left(\frac{\partial \ln \phi}{\partial \theta_k}\right)^2 \phi dX \, dG(\theta) \right\}$$

$$\geq \left\{ \int_{A_s} \int_{R_n} \left[t_k - \theta_k - E\left[\psi_k(\theta)\right] + E(\theta_k)\right] \left(\frac{\partial \ln \phi}{\partial \theta_k}\right) \phi dX \, dG(\theta) \right\}^2 .$$

In view of (1.3.3), (1.3.5), and (1.3.6); (1.3.7) may be written,[1]

$$\mathrm{Var}(t_k-\theta_k)\mathrm{EE}\left[\left(\frac{\partial \ln\phi}{\partial\theta_k}\right)^2 \Big| \theta\right] \geq E^2\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right),$$

and if $\mathrm{EE}[(\partial\ln\phi/\partial\theta_k)^2|\theta] \neq 0$, then

$$\mathrm{Var}(t_k-\theta_k) \geq \frac{E^2\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right)}{E[I_k(\theta)]},$$

where $I_k(\theta) = E[(\partial\ln\phi/\partial\theta_k)^2|\theta]$. Since $\mathrm{Var}(t_k-\theta_k) = \mathrm{EE}[(t_k-\theta_k)^2|\theta] - E^2[\psi_k(\theta)-\theta_k]$, we may write

$$\text{(1.3.8)} \qquad \mathrm{EE}[(t_k-\theta_k)^2|\theta] \geq E^2[\psi_k(\theta)-\theta_k] + \frac{E^2\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right)}{E[I_k(\theta)]}.$$

If $\psi_k(\theta) = \theta_k$, (1.3.8) may be written

$$\text{(1.3.9)} \qquad \mathrm{EE}[(t_k-\theta_k)^2|\theta] \geq \frac{1}{E[I_k(\theta)]}$$

---

[1] EE symbolizes taking the expectation with respect to X for fixed θ and then with respect to θ.

when $\left[\frac{\partial^2\phi}{\partial\theta^2}\right] < K(X)$, where $K(X)$ is integrable over R_n, it is well known that

$$E\left[\left(\frac{\partial \ln\phi}{\partial\theta_k}\right)^2 \Big|\theta\right] = E\left[-\frac{\partial^2 \ln\phi}{\partial\theta_k^2}\Big|\theta\right] .$$

Then we can write $I_k(\theta) = E[-\partial^2 \ln\phi/\partial\theta_k^2|\theta]$ in (1.3.8) and (1.3.9). Since $I_k(\theta)$ is called the amount of information, $E[I_k(\theta)]$ may logically be termed the mean amount of information.

It should be noted that the derivation of these inequalities is equally applicable to samples from multivariate populations.

1.4 The Discrete Case. Suppose that $f(x|\theta)$ is a discrete frequency function whose range, R_1, may be finite or denumerably infinite but independent of θ_k. Assume ϕ is a continuous function of θ_k for all X in R_n and θ in A_s, and that $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} \partial\phi/\partial\theta_k$ and $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} t_k(\partial\phi/\partial\theta_k)$ converge uniformly in A_s. By operations similar to those employed in paragraph 1.3, we find

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} \frac{\partial \ln\phi}{\partial\theta_k}\phi = 0 \tag{1.4.1}$$

and

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} t_k \frac{\partial \ln\phi}{\partial\theta_k}\phi = \frac{\partial\psi_k(\theta)}{\partial\theta_k} \tag{1.4.2}$$

since the assumptions just made allow differentiation under the summation signs. By the Schwarz inequality we may write

$$\left\{ \int_{A_s} \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left[t_k - \theta_k - E\left[\psi_k(\theta)\right] + E(\theta_k)\right]^2 \phi dG(\theta) \right\}$$

$$\cdot \left\{ \int_{A_s} \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left(\frac{\partial \ln \phi}{\partial \theta_k}\right)^2 \phi dG(\theta) \right\}$$

$$\geq \left\{ \int_{A_s} \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left[t_k - \theta_k - E\left[\psi_k(\theta)\right] + E(\theta_k)\right] \frac{\partial \ln \phi}{\partial \theta_k} \phi dG(\theta) \right\}^2 .$$

Following steps analogous to those in paragraph 1.3, we arrive at (1.3.8) and (1.3.9), for the discrete case.

1.5 Conditions for Equality. The condition under which the equalities in (1.3.8) and (1.3.9) hold are set forth in the following three theorems.

Th. 1. If (i) $\Pr. \left\{ E\left[(t_k - \theta_k)^2 | \theta\right] = c_1 \right\} = 1,$

(ii) $\Pr. \left[\psi_k(\theta) = c_2 \right] = 1,$

(iii) $\Pr. \left[\theta_k = c_3 \right] = 1,$

(iv) $\Pr. \left[\dfrac{\partial \psi_k(\theta)}{\partial \theta_k} = c_4 \right] = 1,$ and

(v) $\Pr. \left[L_k(\theta) = c_5 \right] = 1,$

where c_i, $i = 1, 2, \ldots, 5$, are constants, then the equality in (1.3.8) holds if and only if t_k is a sufficient estimate of θ_k.

Proof. Under the conditions of the theorem (1.3.8) reduces to (1.1.1), for which it has been shown by Rao [12] that the equality holds if and only if t_k is a sufficient estimate of θ_k.

Th. 2. If t_k is an unbiased sufficient estimate of θ_k, then the equality in (1.3.9) holds if and only if Pr. $\left[I_k(\theta) = c_5\right] = 1$, where c_5 is a constant.

Proof. Since t_k is an unbiased sufficient estimate of θ_k, we have from Rao [12],

$$E\left[(t_k-\theta_k)^2|\theta\right] = I_k(\theta)^{-1} .$$

Taking expectations with respect to θ, we have

$$EE\left[(t_k-\theta_k)^2|\theta\right] = E\left[I_k(\theta)^{-1}\right] . \tag{1.5.1}$$

Now equality of (1.3.9) requires that,

$$EE\left[(t_k-\theta_k)^2|\theta\right] = \frac{1}{E\left[I_k(\theta)\right]} . \tag{1.5.2}$$

Combining (1.5.1) and (1.5.2), we have

$$E\left[I_k(\theta)^{-1}\right] = \frac{1}{E\left[I_k(\theta)\right]} ,$$

or

$$(1.5.3) \qquad E\left[I_k(\theta)^{-1}\right] E\left[I_k(\theta)\right] = 1.$$

This can be written

$$(1.5.4) \qquad \left\{\int_{A_s} \left[I_k(\theta)\right]^{-1} dG(\theta)\right\} \left\{\int_{A_s} I_k(\theta)\, dG(\theta)\right\} = 1.$$

Now by the Schwarz inequality we have

$$(1.5.5) \qquad \left\{\int_{A_s} I_k(\theta)^{-1}\, dG(\theta)\right\} \left\{\int_{A_s} I_k(\theta)\, dG(\theta)\right\} \geq \left\{\int_{A_s} dG(\theta)\right\}^2 = 1.$$

Obviously when the equality holds in (1.5.5) it is equivalent to (1.5.4). But the equality in (1.5.5) is achieved if and only if for a constant c_5 independent of θ, $c_5 \left[I_k(\theta)\right]^{-1/2} = \left[I_k(\theta)\right]^{1/2}$ with probability one; that is, if and only if $\Pr\left[I_k(\theta) = c_5\right] = 1$, which proves the theorem.

Before proceeding to theorem 3, we cite the following definition.

Definition. Any pair of $\phi(X|\theta)$ and $G(\theta)$ wherein any one of the assumptions (i) - (iv) inclusive of th. 1 does not hold for the θ_k under consideration is termed the non-trivial estimation case.

Th. 3. For the non-trivial estimation case the equality in (1.3.8) is achieved if and only if t_k is an unbiased sufficient estimate of θ_k which is normally distributed with constant variance equal to $I_k(\theta)^{-1}$. Consequently the equality in (1.3.9) is achieved under the same conditions.

Proof. For the non-trivial estimation case the equality in (1.3.7) and consequently in (1.3.8) and (1.3.9) is achieved if and only if there exists a λ independent of X and Θ such that,

$$\lambda \frac{\partial \ln \phi}{\partial \Theta_k} = t_k - \Theta_k - E\left[\psi_k(\Theta)\right] + E(\Theta_k),$$

for almost all X in R_n and Θ in A_s. Integrating, we have

$$\lambda \ln \phi = \Theta_k t_k - \frac{\Theta_k^2}{2} - \Theta_k E\left[\psi_k(\Theta)\right] + \Theta_k E(\Theta_k) + C_1(X, \Theta^*),$$

where $\Theta^* = (\Theta_1, \Theta_2, \ldots, \Theta_{k-1}, \Theta_{k+1}, \ldots, \Theta_s)$. We thus have

$$\phi = C_2(X, \Theta^*) \exp \frac{1}{\lambda} \left\{ \Theta_k t_k - \frac{\Theta_k^2}{2} - \Theta_k E\left[\psi_k(\Theta)\right] + \Theta_k E(\Theta_k) \right\} .$$

This is a special case of the form, found by Pitman [11] and Koopman [10], wherein t_k is a sufficient statistic for Θ_k. Integrating both sides of the above equation over R_n, we have

$$\exp \left\{ \frac{\Theta_k}{\lambda} \left[E(\Theta_k) - E(\psi_k(\Theta)) \right] \right\} \int_{R_n} C_2(X, \Theta^*) \exp(\Theta_k t_k / \lambda) dX = \exp(\Theta_k^2 / 2\lambda).$$

Make the change of variables in the integral,

$$z_i = z_i(X), \qquad i = 1, 2, \ldots, n-1,$$
$$t_k = t_k(X),$$

where $t_k(X)$ and the $z_i(X)$ are unique, continuous, and possess continuous

partial derivatives. Further the transformation is one-to-one. Then we have,

$$\exp\left\{\frac{\theta_k}{\lambda}\left[E(\theta_k)-E\left[\psi_k(\theta)\right]\right]\right\} \int_{B_n} C_3(Z,t_k,\theta^*)\exp(\theta_k t_k/\lambda)\, dZdt_k = \exp(\theta_k^2/2\lambda),$$

where B_n is the range of $(Z,t_k) = (z_1, z_2, \ldots, z_{n-1}, t_k)$. If we integrate out Z, we have

$$(1.5.6) \qquad \int_{b_k} C_4(t_k,\theta)e^{\frac{\theta_k t_k}{\lambda}}\, dt_k = e^{\frac{\theta_k^2}{2\lambda}},$$

where b_k, the range of t_k, may be taken from $-\infty$ to ∞. Then the left hand side of (1.5.6) is a bilateral Laplace transform of $C_4(t_k,\theta)$ with argument θ_k/λ. Recall that θ_k has a non-degenerate range say $\gamma_1 < \theta_k < \gamma_2$. Obviously $e^{\theta_k/2\lambda}$ exists at $\theta_k = \gamma_1 + \epsilon_1$ and $\theta_k = \gamma_2 - \epsilon_2$, where $\epsilon_1, \epsilon_2 > 0$, such that $\epsilon_2 + \epsilon_1 < \gamma_2 - \gamma_1$. Thus we can apply the theorem of Widder [16], p. 238, and conclude that the integral in (1.5.6) converges for θ_k in the vertical strip of the complex plane, $\gamma_1 + \epsilon < \theta_k < \gamma_2 - \epsilon_2$. Thus we can apply the uniqueness theorem of the bilateral Laplace transform (see Widder [16], p. 243) and conclude that $C_4(t_k,\theta) = (1/\sqrt{2\pi\lambda})\exp(-t_k^2/2\lambda)$. Therefore, for equality, the frequency function of t_k must be

$$h(t_k|\theta_k) = \frac{1}{\sqrt{2\pi\lambda}}\, e^{-\frac{(t_k-\theta_k)^2}{2\lambda}}, \quad -\infty < t_k < \infty,$$

where obviously $\lambda = I_k(\theta)^{-1}$. Further the equality holds regardless of the form of the marginal distribution of θ_k.

It should be noted that though theorems 2 and 3 require that $I_k(\theta)$ be a constant, it is not necessary that all the components of θ occurring in $I_k(\theta)$ be constants. It is possible, for instance, that the components of θ occurring in $I_k(\theta)$ have a singular multivariate distribution such that all the probability is located on the hyperplane $I_k(\theta)$ = constant.

Obviously the sample mean from a normal population with constant variance satisfies theorem 3. However, it is by no means the only such estimate. Let

$$f(x|\theta) = \frac{1}{x\sqrt{2\pi c}} e^{-\frac{(\ln x-\theta)^2}{2c}}, \quad 0 < x < \infty \quad ,$$

where c is a constant, which is the so-called logarithmico-normal distribution (see Cramer [6], p. 220). Then, $t = \sum_{i=1}^{n} \ln x_i/n$ is normally distributed with mean θ and variance c/n, which is the minimum variance attainable under the extended inequality.

A situation in which a parameter is assumed to be a random variable is the analysis of variance model II of Eisenhart [8]. The simplest case is the one way classification. Here the model is

$$x_{ij} = \alpha_i + \epsilon_{ij},$$

where x_{ij} is the jth observation in the ith class. We assume there are k classes where the ith class has n_i observations. Suppose α_i and ϵ_{ij}

are random samples of size m and N, $N = \sum_{i=1}^{k} n_i$, from two normally distributed populations having means μ and zero respectively, and variances σ_α^2 and σ_ϵ^2 respectively. Then,

$$\phi = \prod_{i=1}^{m} \prod_{j=1}^{n_i} f(x_{ij}|\alpha_i) = \frac{1}{(\sqrt{2\pi\sigma_\epsilon})^N} \exp\left\{ - \sum_{ij}\sum \frac{(x_{ij}-\alpha_i)^2}{2\sigma_\epsilon^2} \right\},$$

and

$$(1.5.6) \qquad EE\left(- \frac{\partial^2 \ln\phi}{\partial\alpha_i^2} \,|\alpha_i\right) = \frac{n_i}{\sigma_\epsilon^2}, \; i = 1, 2, \ldots, m.$$

The ML estimate for α_i is $\hat{\alpha}_i = \left(\sum_{j=1}^{n_i} x_{ij}\right)/n_i$, $i = 1, 2, \ldots, m$. Here $E(\hat{\alpha}_i|\alpha_i) = \alpha_i$, and $EE[(\hat{\alpha}_i-\alpha_i)^2|\alpha_i] = \sigma_\epsilon^2/n_i$, which by (1.5.6) is the minimum mean square error. Notice the assumption of normality of α_i was not required for equality.

1.6 Most Unfavorable Distributions.[2] In most cases the $G(\Theta)$ is not known, so the lower bound on the mean square error cannot be found. If $\Psi_k(\Theta) = \Theta_k$, it is of interest to know the greatest value the lower bound can attain, as well as the set of $G(\Theta)$ which produces it. To this end define $G_k^*(\Theta)$ to be a most unfavorable distribution with respect to Θ_k if

$$\int_{A_s} I_k(\Theta)\, dG_k^*(\Theta) \leq \int_{A_s} I_k(\Theta)\, dG(\Theta),$$

for all $G(\Theta)$ defined over A_s.

If $I_k(\Theta)$ has a unique minimum with respect to that subset of the

[2] For an analogous concept, least favorable distributions, see Wald [15], p. 18.

parameters appearing therein, then a most unfavorable distribution is one for which the marginal distribution of these parameters is trivial. It may be that $I_k(\theta)$ is independent of all parameters so that all $G(\theta)$ are most unfavorable distributions. A case in point is the Cauchy distribution,

$$f(x|\theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty ,$$

where

$$I(\theta) = \frac{4n}{\pi} \int_{-\infty}^{\infty} \frac{(x-\theta)^2}{[1+(x-\theta)^2]^3} \, dx = \frac{n}{2} .$$

Here $EE[(t-\theta)^2|\theta] \geq 2/n$ regardless of the form of $G(\theta)$. There are also cases in which no most unfavorable distribution exists except possibly when from some prior information A_s is restricted.

1.7 Most Unfavorable Distributions for Some Laplacian Distributions. M. C. K. Tweedie [13, 14] has called a distribution Laplacian if it belongs to the general class of distributions for which the sample mean is a sufficient statistic for one of its parameters. The general form of such a distribution's frequency function is

$$f(x|\theta_1,\theta_2) = e^{-xq(\theta_1)-\theta_2 F(\theta_1)} h(x,\theta_2).$$

That is, of course, a special case of the Pitman-Koopman form (1.1.3). Here we have,

$$E\left[I(\theta_1,\theta_2)\right] = \int_{A_2} \left[m(\theta_1,\theta_2)q''(\theta_1)+\theta_2 F''(\theta_1)\right]g(\theta_1,\theta_2)d\theta_1 d\theta_2,$$

where $E(x|\theta_1,\theta_2) = m(\theta_1,\theta_2)$. If $I(\theta_1,\theta_2)$ has an absolute minimum for some subset A_2' of A_2, and (θ_1^o, θ_2^o) is an element of A_2', then $E\left[I(\theta_1,\theta_2)\right] \geq m(\theta_1^o, \theta_2^o)q''(\theta_1^o) + \theta_2^o F''(\theta_1^o)$, for all (θ_1, θ_2) in A_2. Further, this is the absolute minimum attainable by $E\left[I(\theta_1, \theta_2)\right]$. It is reached when $dG(\theta_1, \theta_2) = 0$, for (θ_1, θ_2) not an element of A_2'. Thus, we have found a set of most unfavorable distributions. This result will now be applied to several specific Laplacian distributions.

Type a. $\theta_2 = 1$. This includes the binomial, Pascal, and Poisson distributions.

(1) Binomial distribution.

$$f(x|\theta_1) = e^{x\ln\left(\frac{\theta}{1-\theta_1}\right)-\ln(1-\theta_1)}, \quad x = 0, 1, \; 0 < \theta_1 < 1,$$

$$q(\theta_1) = -\ln\left(\frac{\theta_1}{1-\theta_1}\right), \; F(\theta_1) = \ln(1-\theta_1), \; m(\theta_1) = \theta_1,$$

and

$$I(\theta_1) = \frac{1}{\theta_1(1-\theta_1)} .$$

$I(\theta_1)$ has a unique absolute minimum at $\theta_1 = 1/2$, so that[3] $G^*(\theta_1) = \epsilon(\theta_1 - 1/2)$ is the only most unfavorable distribution. Further

$$E\left[I(\theta_1)\right] \nless 4N.$$

(2) Pascal distribution.

$$f(x|\theta_1) = e^{x\ln(1-\theta_1) - r\ln(\frac{1-\theta_1}{\theta_1})} \binom{x-1}{r-1}, \quad x = r, r+1, \ldots; \; 0 \leq \theta_1 \leq 1,$$

$$q(\theta_1) = -\ln(1-\theta_1), \; F(\theta_1) = \ln(\frac{1-\theta_1}{\theta_1}), \; m(\theta_1) = r/\theta_1,$$

where r is a known fixed positive integer. $I(\theta_1) = r/\theta_1^2(1-\theta_1)$, which has a unique absolute minimum at $\theta_1 = 2/3$. Therefore $G^*(\theta_1) = \epsilon(\theta_1 - 2/3)$ is the only most unfavorable distribution and $E\left[I(\theta_1)\right] \nless 17N/4$. It should be noted x/r is not an unbiased estimated of θ_1. If we consider $\alpha = 1/\theta_1$ as the parameter to be estimated, then

$$f(x|\alpha) = e^{-x\ln(\frac{\alpha}{\alpha-1}) - \ln(\alpha-1)} \binom{x-1}{r-1} \quad x = r, r+1, \ldots,$$
$$1 < \alpha < \infty,$$

$$q(\alpha) = \ln(\frac{\alpha}{\alpha-1}), \; F(\alpha) = \ln(\alpha-1), \; m(\alpha) = r\alpha.$$

---

[3] Following Cramer [6], p. 192, the distribution function

$$\epsilon(x-a) = \begin{cases} 0 \text{ for } x < a \\ 1 \text{ for } x \geq a. \end{cases}$$

and

$$I(\alpha) = \frac{r}{\alpha(\alpha-1)} .$$

Here x/r is an unbiased estimate of α. However the expression $I(\alpha)$ does not have an absolute minimum in A_1, i.e., $1 \leq \alpha < \infty$, but rather it has a limit of zero as $\alpha \longrightarrow \infty$. Thus letting $\alpha \longrightarrow \infty$ produces a most unfavorable situation, which is equivalent to letting $\theta_1 = 0$. It is interesting to note that though $\theta_1 = 2/3$ was the most unfavorable situation when estimating θ_1, when estimating $\alpha = 1/\theta_1$ we have, in effect, that $\theta_1 = 0$ is the most unfavorable situation. Thus we have established that "most unfavorableness" is not an invariant property.

(3) Poisson distribution.

$$f(x|\theta_1) = e^{x\ln\theta_1 - \theta_1} \frac{1}{x!}, \; x = 0, 1, 2, \ldots; \; 0 < \theta_1 < \infty$$

$$q(\theta_1) = -\ln\theta_1, \; F(\theta_1) = \theta_1, \; m(\theta_1) = \theta_1,$$

$$I(\theta_1) = 1/\theta_1.$$

$I(\theta_1)$ has no absolute minimum in A_1, but rather has a limit of zero as $\theta_1 \longrightarrow \infty$. However, if from some prior consideration we can restrict $\theta_1 < a$, then a most unfavorable distribution is $G^*(\theta_1) = \epsilon(\theta_1 - a)$.

Type b. $\theta_2 \neq 1$, $q(\theta_1) = \theta_1$. Immediately we have $q''(\theta_1) = 0$, and $I(\theta_1, \theta_2) = \theta_2 F''(\theta_1)$.

(1) Gamma distribution.

$$f(x|\theta_1,\theta_2) = e^{-x\theta_1+\theta_2 \ln\theta_1} \left(\frac{x^{\theta_1-1}}{\Gamma(\theta_2)}\right), \quad x > 0,\ \theta_2 > 0,\ \theta_1 > 0.$$

$$F(\theta_1) = -\ln\theta_1,\ I(\theta_1,\ \theta_2) = \frac{\theta_2}{\theta_1^2}.$$

Here $I(\theta_1,\ \theta_2) \longrightarrow 0$, as $\theta_1 \longrightarrow \infty$ and/or $\theta_2 \longrightarrow 0$, but no most unfavorable distribution can be cited unless we assume $\theta_2/\theta_1^2 \leq a$.

(2) Normal distribution (parameters adjusted).

$$f(x|\theta_1,\theta_2) = e^{-\theta_1 x-\theta_2 \frac{\theta_1^2}{2}} \frac{e^{-\frac{x^2}{2\theta_2}}}{2\pi\theta_2}, \quad -\infty < x < \infty,\ \theta_2 > 0,\ -\infty < \theta_1 < \infty,$$

where $\theta_2 = \sigma^2$, $\theta_1 = \frac{-\mu}{\sigma^2}$, in the usual notation.

$$F(\theta_1) = \frac{\theta_1^2}{2} \text{ and } I(\theta_1,\ \theta_2) = \theta_2.$$

Here $\theta_2 \longrightarrow 0$ establishes a minimum, so that if we can restrict $\theta_2 \geq \alpha$, then $\epsilon(\theta_2-a)$ is a most unfavorable distribution.

1.8 More Stringent Inequalities. Bhattacharyya [4] has found greater lower bounds for the mean square errors of estimates in the case

of constant parameters. This admits of direct extension to the present case. We can write (1.3.3) and (1.3.4) respectively as

$$(1.8.1) \qquad E\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\,|\theta\right) = 0$$

and

$$(1.8.2) \qquad \frac{\partial\psi_k(\theta)}{\partial\theta_k} = \mathrm{Cov}\left(t_k-\theta_k, \frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\,|\theta\right).$$

By the result of appendix A,

$$\mathrm{Cov}\left(t_k-\theta_k, \frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\right) = E\left[\mathrm{Cov}\left(t_k-\theta_k, \frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\right)|\theta\right]$$

$$+ \mathrm{Cov}\left[E(t_k-\theta_k|\theta),\ E\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\,|\theta\right)\right].$$

From which by using (1.8.1) and (1.8.2), we obtain

$$(1.8.3) \qquad E\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right) = \mathrm{Cov}\left(t_k-\theta_k, \frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k}\right).$$

With suitable regularity conditions on ϕ and its derivatives similar to those cited in paragraph 1.3, we can differentiate (1.3.2) p times and obtain as in (1.8.3) that

$$(1.8.4) \qquad E\left(\frac{\partial^\beta\psi_k(\theta)}{\partial\theta_k^\beta}\right) = \mathrm{Cov}\left(t_k-\theta_k, \frac{1}{\phi}\frac{\partial^\beta\phi}{\partial\theta_k^\beta}\right), \quad \beta = 1, 2, \ldots, p.$$

Define,

$$J_{\alpha\beta} = \mathrm{Cov}\left(\frac{1}{\phi}\frac{\partial^{\alpha}\phi}{\partial\theta_k^{\alpha}}, \frac{1}{\phi}\frac{\partial^{\beta}\phi}{\partial\theta_k^{\beta}}\right), \ \alpha, \beta = 1, 2, \ldots, p.$$

Let $J = \left[J_{\alpha\beta}\right]$ and $J^{-1} = \left[J^{\alpha\beta}\right]$. Denote by $R_{0\cdot 12\ldots p}$ the multiple correlation coefficient between $t_k - \theta_k$ and $(1/\phi)\ (\partial\phi/\partial\theta_k)$, $(1/\phi)\ (\partial^2\phi/\partial\theta_k^2)$, ..., $(1/\phi)\ (\partial^p\phi/\partial\theta_k^p)$. Then by a result cited by Wilks [17], p. 42 ff.,

$$R^2_{0.123\ldots p} = \frac{\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^{\alpha}\psi_k(\theta)}{\partial\theta_k^{\alpha}}\right) E\left(\frac{\partial^{\beta}\psi_k(\theta)}{\partial\theta_k^{\beta}}\right) J^{\alpha\beta}}{\mathrm{Var}(t_k - \theta_k)}$$

Since $R^2_{0.123\ldots p} \leq 1$, we may write

$$\mathrm{Var}(t_k - \theta_k) \geq \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^{\alpha}\psi_k(\theta)}{\partial\theta_k^{\alpha}}\right) E\left(\frac{\partial^{\beta}\psi_k(\theta)}{\partial\theta_k^{\beta}}\right) J^{\alpha\beta},$$

from which we have

$$\text{(1.8.5)} \qquad EE\left[(t_k - \theta_k)^2 | \theta)\right] \geq E^2(\psi_k(\theta) - \theta_k) +$$

$$\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^{\alpha}\psi_k(\theta)}{\partial\theta_k^{\alpha}}\right) E\left(\frac{\partial^{\beta}\psi_k(\theta)}{\partial\theta_k^{\beta}}\right) J^{\alpha\beta}.$$

This is a greater lower bound than that of (1.3.8) since the multiple correlation between $t_k - \theta_k$ and the above series of variates will be larger than the simple correlation between $t_k - \theta_k$ and $(1/\phi)\ (\partial\phi/\partial\theta_k)$. This latter correlation is essentially what was used in deriving (1.3.8). It should be noted that this method of obtaining a higher lower bound applies only if $\psi_k(\theta)$ is non-linear function of θ_k and consequently is not applicable in the unbiased case. As noted in Wilks [17], p. 46, the equality holds if and only if all the probability in the p + 1 dimensional space of the random variables lies on the surface,

$$t_k-\theta_k-E\left[\psi_k(\theta)\right] + E(\theta_k) = \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \frac{1}{\phi} \frac{\partial^{\alpha}\phi}{\partial\theta_k^{\alpha}} \cdot E\left(\frac{\partial^{\beta}\psi_k(\theta)}{\partial\theta_k^{\beta}}\right) J^{\alpha\beta} .$$

1.9 The Sequential Case. Wolfowitz [18] has extended the Cramer-Rao Inequality to situations where the sample size is a random variable depending on the sequence of observations. In our notation his result is

$$E(t_k-\theta_k)^2 \geq \left[\psi_k(\theta)-\theta_k\right]^2 + \frac{\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right)^2}{E(n|\theta)E\left[\left(\frac{\partial\ln f(x|\theta)}{\partial\theta_k}\right)^2 |\theta\right]} .$$

We shall proceed to extend this result to the case where θ is a random variable.

Under suitable regularity conditions Wolfowitz has shown,

$$E\left(\frac{\partial \ln \phi}{\partial \theta_k} \,|\theta\right) = 0,$$

and

$$(1.9.1) \qquad E\left[\left(\frac{\partial \ln \phi}{\partial \theta_k}\right)^2 |\theta\right] = E(n|\theta)E\left[\left(\frac{\partial \ln f(x|\theta)}{\partial \theta_k}\right)^2 |\theta\right] .$$

By definition,

$$\sum_{j=1}^{\infty} \int_{R_j} t_k(x_1, x_2, \ldots, x_j) \prod_{i=1}^{j} f(x_i|\theta)dx_i = \psi_k(\theta).$$

Under the regularity conditions cited by Wolfowitz we may differentiate under the integral sign and obtain

$$E(t_k \cdot \frac{\partial \ln \phi}{\partial \theta_k} \,|\theta) = \text{Cov}\,(t_k - \theta_k, \frac{\partial \ln \phi}{\partial \theta_k} \,|\theta) = \frac{\partial \psi_k(\theta)}{\partial \theta_k} .$$

The result of appendix A yields,

$$E\left[\frac{\partial \psi_k(\theta)}{\partial \theta_k}\right] = \text{Cov}(t_k - \theta_k, \frac{\partial \ln \phi}{\partial \theta_k}) .$$

Since the square of the correlation coefficient of any two variates cannot exceed unity, we have,

$$E^2\left(\frac{\partial \psi_k(\theta)}{\partial \theta_k}\right) \leq \text{Var}\left(\frac{\partial \ln \phi}{\partial \theta_k}\right) . \text{Var}(t_k - \theta_k),$$

and

$$\mathrm{Var}(t_k-\theta_k) \geq \frac{E^2\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right)}{E\left\{E(n|\theta)E\left[\left(\frac{\partial \ln f(x|\theta)}{\partial\theta_k}\right)^2|\theta\right]\right\}} .$$

This may be written

$$(1.9.2) \qquad EE\left[(t_k-\theta_k)^2|\theta\right] \geq E^2\left[\psi_k(\theta)-\theta_k\right] + \frac{E^2\left(\frac{\partial\psi_k(\theta)}{\partial\theta_k}\right)}{E\left\{E(n|\theta)\cdot E\left[\left(\frac{\partial \ln f(x|\theta)}{\partial\theta_k}\right)^2|\theta\right]\right\}} .$$

When $\psi_k(\theta) = \theta_k$, (1.9.2) becomes,

$$(1.9.3) \qquad EE\left[(t_k-\theta_k)^2|\theta\right] \geq \frac{1}{E\left\{E(n|\theta)E\left[\left(\frac{\partial \ln f(x|\theta)}{\partial\theta_k}\right)^2|\theta\right]\right\}} .$$

These results are valid for discrete as well as continuous distributions.

A simple example of sequential estimation involves sampling from a binomial population until a specified number of successes, say r, occur. Here $f(x|\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, and $E(n|\theta) = r/\theta$. Therefore, for $\psi(\theta) = \theta$,

$$EE\left[(t-\theta)^2|\theta\right] \geq \frac{1}{rE\left(\frac{1}{\theta^2(1-\theta)}\right)} .$$

This result corresponds exactly to that obtained in paragraph 1.7 for the Pascal distribution.

1.10 Linear Estimation of $E(\theta_k)$. Consider m samples $X_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})$, $i = 1, 2, \ldots m$, chosen from a population $f(x|{}^i\theta)$, which are randomly and independently chosen from a super-population of populations with frequency functions of the form $f(x|\theta)$. Thus for each sample X_i, there is associated an unobserved random variable ${}^i\theta$, $i = 1, 2, \ldots, m$, with distribution $G({}^i\theta)$. We seek to find an estimate of $E(\theta_k)$, say T_k, where $1 \leq k \leq s$. It is supposed that for each sample there exists an unbiased estimate of ${}^i\theta_k$, namely ${}^it_k$, and we restrict our discussion to the set of T_k which are linear functions of the ${}^it_k$, that is,

$$T_k = \sum_{i=1}^{m} c_i \; {}^it_k$$

where $(c_1, c_2, \ldots, c_m)$ is a vector of real numbers. If we further restrict ourselves to unbiased estimates of $E(\theta_k)$, it follows that $\sum_{i=1}^{m} c_i = 1$.

The minimum variance unbiased estimate of $E(\theta_k)$, $\hat{T}_k$, is found by minimizing the expression,

$$\mathrm{Var}(T_k) = \sum_{i=1}^{m} c_i^2 \; \mathrm{Var}({}^it_k),$$

with respect to the c's, subject to the restriction $\sum_{i=1}^{m} c_i = 1$. This yields the normal equations:

$$(1.10.1) \qquad \hat{c}_i \ \mathrm{Var}({}^i t_k) + \lambda = 0, \ i = 1, 2, \ldots, m$$

$$\sum_{i=1}^{m} \hat{c}_i = 1,$$

where λ is a Lagrangian multiplier.

Consider now the variance of the minimum estimate $\hat{T}_k$, found by solving (1.10.1). We have

$$(1.10.2) \qquad \mathrm{Var}(\hat{T}_k) = \sum_{i=1}^{m} \hat{c}_i^2 \ \mathrm{Var}({}^i t_k).$$

By the result of appendix A,

$$\mathrm{Var}({}^i t_k) = E\left[\mathrm{Var}({}^i t_k | \theta) \right] + \mathrm{Var}\left[E({}^i t_k | \theta) \right],$$

from which it follows that,

$$(1.10.3) \qquad \mathrm{Var}({}^i t_k) = EE\left[({}^i t_k - {}^i \theta_k)^2 / \theta \right] + \mathrm{Var}({}^i \theta_k).$$

So that (1.10.2) may be written

$$(1.10.4) \qquad \mathrm{Var}(\hat{T}_k) = \sum_{i=1}^{m} \hat{c}_i^2 \ EE\left[({}^i t_k - {}^i \theta_k)^2 | \theta \right] + \mathrm{Var}(\theta_k) \sum_{i=1}^{m} \hat{c}_i^2 .$$

Applying (1.3.9) to (1.10.4), we have

$$(1.10.5) \qquad \mathrm{Var}(\hat{T}_k) \geq \sum_{i=1}^{m} \frac{\hat{c}_i^2}{EE\left[\left(\frac{\partial \ln \phi(X_i | \theta)}{\partial \theta_k} \right)^2 \Big| \theta \right]} + \mathrm{Var}(\theta_k) \sum_{i=1}^{m} \hat{c}_i^2 ,$$

the equality being achieved under the conditions cited in paragraph 1.5.

We may apply these results to the analysis of variance model II cited in paragraph 1.5. To simplify the normal equations above, let $n_i = n$, $i = 1, 2, \ldots, m$. In the notation of this paragraph,

$$^{i}\theta_k = \alpha_i,$$

$$^{i}t_k = \hat{\alpha}_i = \left(\sum_{j=1}^{n} x_{ij} \right)/n, \quad i = 1, 2, \ldots, m,$$

and (1.10.1) becomes

$$\hat{c}_i \left(\frac{\sigma_\epsilon^2}{n} + \sigma_\alpha^2 \right) + \lambda = 0,$$

$$\sum_{i=1}^{m} \hat{c}_i = 1.$$

Solving, we have $\hat{c}_i = 1/m$, $i = 1, 2, \ldots, m$. Thus

$$\hat{T}_k = \left(\sum_{i=1}^{m} \hat{\alpha}_i \right)/m = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \right)/mn,$$

and

$$\mathrm{Var}\,(\hat{T}_k) = \frac{1}{m} \left(\frac{\sigma_\epsilon^2}{n} + \sigma_\alpha^2 \right),$$

which equals the lower bound given by (1.10.5).

II. THE ASYMPTOTIC PROPERTIES OF ML ESTIMATORS WHEN SAMPLING FROM ASSOCIATED POPULATIONS

2.1 Summary. Cramer [24], p. 498, points out that the method of maximum likelihood was proposed as a general method of estimation by Fisher [27] in 1912. However, rigorous proofs of the asymptotic properties of the maximum likelihood estimators (hereafter called ML estimators) were not derived until the work of Dugue [25], Wald [32, 33], Cramer [24], Huzurbazar [28], and Chanda [21], all of whom published within the last twenty-five years. Each of these authors assumed observations from a single population. Neyman and Scott [30] and Kraft and LeCam [29] have pointed out situations wherein the ML estimator is not consistent. Their examples involved sampling from a number of subpopulations with different distributions but with only a finite number of observations from each of the subpopulations.

Some situations arise where observations do not come from a single population but from distinct but related populations, related in the sense that some populations reasonably may be assumed to have some parameters in common. We call such populations <u>associated</u>. Such situations may arise, for example, in the combination of results from several experiments. The theory developed may also assist in certain problems of mathematical statistics--it assists in developing the theory for a generalization of a method of paried comparisons indicated briefly in the final section of this paper.

In this paper we extend the basic theory on ML estimation to associated populations obtaining asymptotic results when sample sizes for

for the associated populations become large in constant ratios. Under regularity conditions, the ML estimators for parameters in associated populations are shown to be consistent and asymptotically normal with a variance-covariance matrix as derived. The asymptotic χ^2 - distribution of $-2 \ln \lambda$, λ being the likelihood ratio, holds and, for the non-null distributions, the parameter of noncentrality is set forth.

We limit consideration to ML estimators that are roots of the normal equations for the maximization process. The consistency of the more generally defined estimator can be proved for associated populations following the approach of Wald [33]. Our proof of consistency follows that of Chanda closely; most of the results for associated populations are easy generalizations of results for a single population and results that do not differ essentially from those demonstrations are cited without proof.

2.2 Notation and Assumptions. Let $f_i(x_i, \Theta)$ $i = 1, \ldots, m$ denote density or probability functions (discrete or continuous) where x_i is a random vector with values over a region R_i independent of $\Theta = (\Theta_1, \ldots, \Theta_k)$, an unknown parameter vector lying in a k-dimensional parameter space Ω . $f_i(x_i, \Theta)$ may not depend on all of $\Theta_1, \ldots, \Theta_k$. Let $x_{i1}, \ldots x_{in_i}$ be n_i independent observation vectors on x_i. The joint likelihood function is then

$$\phi = \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} f_i(x_{i\alpha}, \Theta) \tag{2.2.1}$$

and the normal equations are

$$\frac{\partial \ln \phi}{\partial \Theta_r} = 0, \quad r = 1, \ldots, k \tag{2.2.2}$$

from the maximization of $\ln\phi$ and hence of ϕ.

We make the following assumptions:

(1) For almost all $x_i \epsilon R_i$ and for all $\Theta \epsilon \Omega$,

$$\frac{\partial \ln f_i}{\partial \Theta_r}, \quad \frac{\partial^2 \ln f_i}{\partial \Theta_r \partial \Theta_s} \quad \text{and} \quad \frac{\partial^3 \ln f_i}{\partial \Theta_r \partial \Theta_s \partial \Theta_t}$$

exist for r, s, t = 1, ..., k and i = 1, ..., m.

(2) For all f_i which are densities for almost all $x_i \epsilon R_i$ and for every $\Theta \epsilon \Omega$,

$$(2.2.3) \quad \left|\frac{\partial f_i}{\partial \Theta_r}\right| < F_{ir}(x_i), \quad \left|\frac{\partial^2 f_i}{\partial \Theta_r \partial \Theta_s}\right| < F_{irs}(x_i) \text{ and } \left|\frac{\partial^3 \ln f_i}{\partial \Theta_r \partial \Theta_s \partial \Theta_t}\right| < H_{irst}(x_i)$$

where $F_{ir}(x_i)$ and $F_{irs}(x_i)$ are integrable over R_i and

$$(2.2.4) \qquad \int_{R_i} H_{irst}(x_i)\, dx_i < M_i$$

i = 1, ..., m; r, s, t = 1, ..., k, for all $\Theta \epsilon \Omega$ and where the M_i are finite positive constants. Correspondingly, for all f_i which are probability functions,

$$\sum_{x_i \epsilon R_i} \frac{\partial f_i}{\partial \Theta_r} \quad \text{and} \quad \sum_{x_i \epsilon R_i} \frac{\partial^2 f_i}{\partial \Theta_r \partial \Theta_s}$$

converge uniformly for all $\Theta \epsilon \Omega$ and

$$\left| \frac{\partial^3 \ln f_i}{\partial\theta_r \partial\theta_s \partial\theta_t} \right| < H_{irst}(x_i)$$

where

$$\sum_{x_i \in R_i} H_{irst}(x_i) < M_i$$

for all $\theta \in \Omega$, r, s, t = 1, ..., k; i = 1, ..., m and where the M_i are finite positive constants. These assumptions will permit the interchange of order of differentiation and integration or summation.

(3) For all $\theta \in \Omega$, the matrix $J = \left[J_{rs}(\theta) \right]$ with

$$(2.2.5) \qquad J_{rs}(\theta) = \sum_{i=1}^{m} \mu_i \int_{R_i} \frac{\partial \ln f_i}{\partial\theta_r} \frac{\partial \ln f_i}{\partial\theta_s} f_i \, dx_i$$

is positive definite with finite determinant. [Replace $\int_{R_i}$ by $\sum_{x_i \in R_i}$ for those f_i which are probability functions in (2.2.5)]. We define $\mu_i = n_i/N$ where $N = \sum_{i=1}^{m} n_i$.

2.3 Consistency. Let $\theta^o \in \Omega$ be the unknown true value of θ. Consider the following Taylor's expansion:

$$(2.3.1) \qquad \frac{\partial \ln f_i}{\partial\theta_r} = \left. \frac{\partial \ln f_i}{\partial\theta_r} \right|_{\theta = \theta^o} + \sum_{s=1}^{k} (\theta_s - \theta_s^o) \left. \frac{\partial^2 \ln f_i}{\partial\theta_r \partial\theta_s} \right|_{\theta = \theta^o}$$

$$+ \frac{1}{2} \sum_{s,t=1}^{k} (\theta_s - \theta_s^o)(\theta_t - \theta_t^o) \left. \frac{\partial^3 \ln f_i}{\partial\theta_r \partial\theta_s \partial\theta_t} \right|_{\theta = \theta'}$$

where $\theta' = \theta'(x_i)$ depends on x_i and is such that it lies within the hypercell of which $\theta - \theta^o$ is the diagonal for all x_i. Summing over all observations of all subpopulations, we see that (2.3.1) implies

$$(2.3.2) \quad L_r(\theta) = L_r(\theta^o) - \sum_{s=1}^{k} (\theta_s-\theta_s^o)L_{rs}(\theta^o) + \frac{1}{2} \sum_{s,t=1}^{k} (\theta_s-\theta_s^o)(\theta_t-\theta_t^o)L_{rst}$$

where

$$L_r(\theta) = \frac{1}{N} \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r} = \sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r},$$

$$L_{rs}(\theta) = -\frac{1}{N} \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} \frac{\partial^2 \ln f_{i\alpha}}{\partial \theta_r \partial \theta_s} = \sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \left(-\frac{1}{n_i}\right) \frac{\partial^2 \ln f_{i\alpha}}{\partial \theta_r \partial \theta_s}$$

and

$$L_{rst} = \frac{1}{N} \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} \left. \frac{\partial^3 \ln f_{i\alpha}}{\partial \theta_r \partial \theta_s \partial \theta_t} \right|_{\theta = \theta'}$$

$$= \sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} \left. \frac{\partial^3 \ln f_{i\alpha}}{\partial \theta_r \partial \theta_s \partial \theta_t} \right|_{\theta = \theta'}$$

and $f_i(x_{i\alpha}) = f_{i\alpha}$.

In view of our assumptions, we have

(1) $\quad E\left(\frac{\partial \ln f_i}{\partial \theta_r}\right) = 0, \; r = 1, \ldots, k; \; i = 1, \ldots, m.$

(2) $\quad \sum_{i=1}^{m} \mu_i \, E\left(-\frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s}\right) = J_{rs}(\theta), \; r, s = 1, \ldots, k.$

and

$$(3) \quad E\left(\left|\frac{\partial^3 \ln f_i}{\partial\theta_r \partial\theta_s \partial\theta_t}\right|\right) < M_i, \; r, s, t = 1, \ldots, k; \quad i = 1, \ldots, m,$$

for all $\theta \epsilon \Omega$. Also, from (3), it follows that

$$|L_{rst}| \leq \sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} \left| \frac{\partial^3 \ln f_{i\alpha}}{\partial\theta_r \partial\theta_s \partial\theta_t} \right|_{\theta = \theta'} < \sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} H_{irst}(x_{i\alpha}).$$

Further,

$$\sum_{i=1}^{m} \mu_i \, E| H_{irst}(x_{i\alpha}) | < \sum_{i=1}^{m} \mu_i M_i .$$

If we let $n_i \longrightarrow \infty$, $i = 1, \ldots, m$ so that $n_i/N = \mu_i$, then by Khintchine's Theorem [6], p. 253,

$$(1) \quad \frac{1}{n_i} \sum_{\alpha=1}^{n_i} \frac{\partial \ln f_{i\alpha}}{\partial\theta_r} \Bigg|_{\theta = \theta^o} \xrightarrow[p]{} 0,$$

$$(2) \quad \frac{1}{n_i} \sum_{\alpha=1}^{n_i} \left(- \frac{\partial^2 \ln f_{i\alpha}}{\partial\theta_r \partial\theta_s}\right) \Bigg|_{\theta = \theta^o} \xrightarrow[p]{} E\left(- \frac{\partial^2 \ln f_i}{\partial\theta_r \partial\theta_s}\right) \Bigg|_{\theta = \theta^o},$$

and

$$(3)\quad \frac{1}{n_i}\sum_{\alpha=1}^{n_i}\left|\frac{\partial^3 \ln f_{i\alpha}}{\partial\theta_r\partial\theta_s\partial\theta_t}\right|\Bigg|_{\theta=\theta'} < \frac{1}{n_i}\sum_{\alpha=1}^{n_i} H_{irst}(x_{i\alpha}) \xrightarrow[p]{} E\left[H_{irst}(x_{i\alpha})\right] < M_i$$

for $i = 1, \ldots, m$. Using Slutsky's Theorem [24], p. 255, we have

$$(1)\quad L_r(\theta^o) \xrightarrow[p]{} 0,$$

$$(2)\quad L_{rs}(\theta^o) \xrightarrow[p]{} J_{rs}(\theta^o), \text{ and}$$

$$(3)\quad |L_{rst}| < \sum_{i=1}^{m} \mu_i M_i \text{ as a limiting bound in probability.}$$

In other words, given η and ϵ, two arbitrarily chosen small positive constants and R, a finite positive constant greater than $\sum_{i=1}^{m} \mu_i M_i$, there exists $n_{oi} = n_{oi}(\eta, \epsilon)$ such that, for all $n_i > n_{oi}$ with $n_i = \mu_i \sum_{i=1}^{m} n_i$, $i = 1, \ldots, m$.

$$(2.3.3)\quad P\left[\, |L_r(\theta^o)| < \eta,\ |L_{rs}(\theta^o) - J_{rs}(\theta^o)| < \eta,\ |L_{rst}| < R \,\right] > 1 - \epsilon.$$

The likelihood equations for the estimation of θ are given by putting $L_r(\theta) = 0$, $r = 1, \ldots, k$, in (2.3.2). Thus if $\hat{\theta}$ is the ML estimator of θ, from (2.3.2),

$$(2.3.4)\quad \sum_{s=1}^{k} (\hat{\theta}_s - \theta_s^o)\, L_{rs}(\theta^o) = L_r(\theta^o) + \frac{1}{2}\sum_{s,t=1}^{k} (\hat{\theta}_s - \theta_s^o)(\hat{\theta}_t - \theta_t^o)\, L_{rst}\,,$$

$$r = 1, \ldots, k.$$

By assumption (3), J^{-1} exists for all $\theta \epsilon \Omega$ and, since $\left[L_{rs}(\theta^o) \right] \xrightarrow[p]{} J_{rs}(\theta^o)$, for sufficiently large N, $n_i = \mu_i N$, the inverse matrix $\left[L_{rs}(\theta^o) \right]^{-1}$ exists also. Accordingly, from (2.3.4),

$$(2.3.5) \quad (\hat{\theta}_r - \theta_r^o) = \sum_{p=1}^{k} L_p(\theta^o) L^{pr}(\theta^o) + \frac{1}{2} \sum_{s,t=1}^{k} (\hat{\theta}_s - \theta_s^o)(\hat{\theta}_t - \theta_t^o) \left[\sum_{p=1}^{k} L_{pst} L^{pr}(\theta^o) \right],$$

$$r = 1, \ldots, k,$$

where $\left[L^{rs}(\theta^o) \right] = \left[L_{rs}(\theta^o) \right]^{-1}$, the latter matrix being one of finite elements independent of N. Given η and ϵ positive and arbitrarily small and $T > R$ finite, we can choose $n_{oi}(\eta, \epsilon, T)$ such that, for all $n_i > n_{oi}$, $n_i = \mu_i N$, $i = 1, \ldots, m$,

$$(2.3.6) \quad P\Bigg[\left| \sum_{p=1}^{k} L_p(\theta^o) L^{pr}(\theta^o) \right| \leq \sum_{p=1}^{k} \left| L_p(\theta^o) \right| \left| L^{pr}(\theta^o) \right| < \eta,$$

$$\left| \sum_{p=1}^{k} L_{pst} L^{pr}(\theta^o) \right| \leq \sum_{p=1}^{k} \left| L_{pst} \right| \left| L^{pr}(\theta^o) \right| < T \text{ for all } r \Bigg] > 1 - \epsilon.$$

Obviously for $n_i > n_{io}$, the equations (2.3.5) admit a solution $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$ such that $(\hat{\theta}_r - \theta_r^o)$, $r = 1, \ldots, k$, are of the same order as η, the second term on the right of (2.3.5) being of smaller order than η. Thus,

$$(2.3.7) \qquad P\left[|\hat{\theta}_1 - \theta_1^o| < \eta, \ldots, |\hat{\theta}_r - \theta_r^o| < \eta \right] > 1 - \epsilon$$

for $n_i > n_{io}$, $i = 1, \ldots, m$. It follows that there exists at least one

solution $\hat{\theta}$ of the likelihood equations which is a consistent estimator of the vector θ^o. That the solution is unique may be proved following closely on a proof given by Chanda.

We now state two theorems without proof indicating only necessary modifications of Chanda's proofs.

Theorem 1. The matrix $\left[\left. \frac{\partial^2 \ln \phi}{\partial \theta_r \partial \theta_s} \right|_{\theta = \hat{\theta}} \right]$, where $\hat{\theta}$ is a consistent estimator of θ^o and a solution of the likelihood equations, is negative definite with probability approaching unity as $n_i = \mu_i N \longrightarrow \infty$. The proof follows Chanda [21], p. 59-60. The only changes required are that now

$$J_{rs}(\theta) = \sum_{i=1}^{m} \mu_i \left(- \frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s} \right) ,$$

and we require use of both Khintchine's Theorem and Slutsky's Theorem where Chanda required the former.

Theorem 2. Of all possible solutions to equations (2.3.5), one and only one tends in probability to the true parameter vector θ^o. The proof of Theorem 2 parallels that of Chanda [21], p. 60-61.

We have now seen that the usual consistency properites of ML estimators carry over to estimation when sampling from associated populations.

2.4 Asymptotic Normality. We rewrite (2.3.4) in the form

$$\sum_s (\hat{\theta}_s - \theta_s^o) L_{rs}(\theta^o) - \frac{1}{2} \sum_{s,t=1}^{k} (\hat{\theta}_s - \theta_s^o)(\hat{\theta}_t - \theta_t^o) L_{rst} = L_r(\theta^o), \; r = 1, \ldots, k,$$

and in matrix notation

$$(2.4.1) \qquad (L + G)\,(\hat{\theta}' - \theta') = \left[\sum_{i=1} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r} \bigg|_{\theta = \theta^o} \right]'$$

where $\hat{\theta}'$, θ', and the right-member of (2.4.1) are k-element column vectors, the latter based on the definition of $L_r(\theta^o)$ following (2.3.2), and where $L = [\, L_{rs}(\theta^o) \,]$ and

$$G = \left[-\frac{1}{2} \sum_{t=1}^{k} (\hat{\theta}_t - \theta_t^o) L_{rst} \right] ,$$

L and G being k-square symmetric matrices. We have seen that $L \xrightarrow[p]{} J_o$, $J_o = J\Big|_{\theta = \theta^o}$, positive definite, and $G \xrightarrow[p]{} 0$ (since $\hat{\theta} \xrightarrow[p]{} \theta^o$ and $|L_{rst}|$ is bounded). Hence, for large n_i, we may invert L + G and write

$$\hat{\theta}' - \theta^{o\prime} = (L + G)^{-1} \left[\sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{n_i} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r} \bigg|_{\theta = \theta^o} \right]'$$

$$(2.4.2) \quad \sqrt{N}\,(\hat{\theta}' - \theta^{o\prime}) = (L + G)^{-1} \left[\sum_{i=1}^{m} \mu_i \sum_{\alpha=1}^{n_i} \frac{1}{\sqrt{n_i}} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r} \bigg|_{\theta = \theta^o} \right]' .$$

Since $L \xrightarrow[p]{} J_o$ and $G \xrightarrow[p]{} 0$, by Slutsky's Theorem, $L + G \xrightarrow[p]{} J_o$ and $(L + G)^{-1} \xrightarrow[p]{} J_o^{-1}$. Further, by the multivariate form of the Central Limit Theorem [23], p. 113,

$$\left[\sum_{i=1}^{m} \sqrt{\mu_i} \sum_{\alpha=1}^{n_i} \frac{1}{\sqrt{n_i}} \frac{\partial \ln f_{i\alpha}}{\partial \theta_r} \bigg|_{\theta = \theta^o} \right]'$$

has a multivariate normal distribution with means zero and variance-covariance matrix

$$\left[\sum_{i=1}^{m} \mu_i \, E\left(\frac{\partial \ln f_i}{\partial \theta_r} \cdot \frac{\partial \ln f_i}{\partial \theta_s} \right) \bigg|_{\theta = \theta^o} \right] = - \left[\sum_{i=1}^{m} \mu_i \, E\left(\frac{\partial^2 \ln f_i}{\partial \theta_r \partial \theta_s} \right) \bigg|_{\theta = \theta^o} \right] = J_o \, .$$

Applying a Theorem of Chiang [22], p. 338, we may state that the asymptotic distribution of $\sqrt{N}\,(\hat{\theta}' - \theta^{o\prime})$ is multivariate normal with means zero and variance-covariance matrix $J_o^{-1} J_o J_o^{-1} = J_o^{-1}$ as $n_i = \mu_i N$, $i = 1, \ldots, m$, and $N \longrightarrow \infty$.

2.5 Asymptotic Distribution of the Likelihood Ratio. Let ω be a sub-space of Ω defining an hypothesis H_ω through the relations,

$$(2.5.1) \qquad \xi^1(\theta) = \xi^2(\theta) = \ldots = \xi^r(\theta) = 0, \; r < k.$$

We require the functions in (2.5.1) to be such that:

(a) There exist k-r additional functions $\xi^{r+1}(\theta), \ldots, \xi^k(\theta)$ so that the inverse relationships $\theta_1(\xi), \ldots, \theta_k(\xi)$ exist.

(b) The first and second order partial derivatives of $\xi^1(\theta), \ldots, \xi^k(\theta)$ are uniformly continuous and bounded functions of θ.

(c) The greatest lower bound of the absolute value of the Jacobean $\partial(\xi^1, \ldots, \xi^k)/\partial(\theta_1, \ldots, \theta_k)$ is positive.

The definitions of ω and H_ω conforms with those of Wald [32], p. 463. Wald considered the distribution of $\lambda_N(\omega, E_N)$, the likelihood ratio statistic for comparing ω with Ω and wherein E_N denoted the observed point in the sample space. Wald assumed that the likelihood ratio test is uniformly consistent and stated that this would be proved in a later paper based on weak assumptions on the density function $f(x, \theta)$. (The later proof seems not to have been published.) Based on his assumptions, Wald proved that

$$(2.5.2) \qquad \lim_{N \longrightarrow \infty} \left\{ P\left[-2\ln \lambda_N(\omega, E_N) < t \mid \theta \right] - F_r\left[\lambda_N^2(\theta), t \right] \right\} = 0$$

uniformly in t and θ where $F_r\left[\lambda_N^2(\theta), t\right]$ is the non-central χ^2 - distribution with r degrees of freedom and parameter of non-centrality

$$(2.5.3) \qquad \lambda_N^2(\theta) = N \sum_{p=1}^{r} \sum_{q=1}^{r} \xi^p(\theta) \; \xi^q(\theta) \bar{c}_{pq}^{\,*}(\theta).$$

In (2.5.3), $\left[\bar{c}_{pq}^{\,*}(\theta) \right] = \left[\sum_{l=1}^{k} \sum_{m=1}^{k} \frac{\partial \xi^p}{\partial \theta_l} \frac{\partial \xi^q}{\partial \theta_m} \sigma_{lm}(\theta) \right]$, $p, q = 1, \ldots, r$, with

$$\left[\sigma_{ij}(\theta) \right] = \left[c_{ij}(\theta) \right]^{-1} \text{ and } c_{ij}(\theta) = -E\left(\frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_j} \right), \; i, j = 1, \ldots, k.$$

It was our original intent to generalize Wald's proofs to obtain the analogous results for associated populations. We do not believe however that Wald's conditions are in a form readily verified for specific applications and utilizations of his results. The extensions of theorems on maximum

likelihood procedures to associated populations do seem straightforward and the results corresponding to those of Wald should follow quite directly. We should then obtain, for ω and H_ω defined by (2.5.1), for associated populations that (2.5.2) holds as stated where now only

$$c_{ij}(\theta) = J_{ij}(\theta) = \sum_{h=1}^{m} \mu_h \, E\left(\frac{-\partial^2 \ln f_h}{\partial\theta_i \partial\theta_j}\right)$$

developed in previous sections thus redefining $\lambda_N^2(\theta)$ in (2.5.3). The central χ^2 - distribution would of course be implied by (2.5.2) when θ for which H_ω is true.

The asymptotic power functions of test procedures are often used in comparisons of their properties and one commonly used comparison of two tests results in the evaluation of their asymptotic relative efficiency. To obtain an asymptotic relative efficiency, <u>local</u> alternatives to a null hypothesis, e.g., [31] are considered. In this way, let θ^ω denote a value of $\theta \epsilon \omega$ and let $\left\{ \theta^N \right\}$ be a sequence of values of θ such that $\theta_i^N = \theta_i^\omega + \delta_{iN}(\omega)/N$ with $\lim_{N \to \infty} \delta_{iN}(\omega) = \delta_i(\omega)$ $i = 1, \ldots, k$. Under these conditions,

$$\lambda_N^2(\theta) \longrightarrow \lambda^2(\delta) = \sum_{p=1}^{r} \sum_{q=1}^{r} \sum_{i=1}^{k} \sum_{j=1}^{k} \delta_i \delta_j \left.\frac{\partial \xi^p(\theta)}{\partial\theta_i}\right|_{\theta = \theta^\omega} \cdot \left.\frac{\partial \xi^q(\theta)}{\partial\theta_j}\right|_{\theta = \theta^\omega} \cdot \bar{c}^{*}_{pq}(\theta^\omega)$$

and

$$F_r\left[\lambda_N^2(\theta), t\right] \longrightarrow F_r\left[\lambda^2(\delta), t\right] .$$

From the fact that (2.5.2) holds uniformly in θ, it follows that

$$\lim_{N \longrightarrow \infty} \left\{ P\left[-2\ln \lambda_N^2(\omega, E_N) < t \,|\, \theta \right] - F_r\left[\lambda^2(\delta),\ t\right] \right\} = 0$$

uniformly in t and given the sequence $\left\{ \theta^N \right\}$.

2.6 An Application. Bradley and Terry [20] and Bradley [19] have devised a method of paired comparisons and investigated some of its properties. Positive parameters $\pi_i, \ldots, \pi_t$, $\sum_{i=1}^{t} \pi_i = 1$ were postulated to represent the effects of treatments $T_1, \ldots, T_t$ and to operate so that, if X_i is a response to T_i,

$$P(X_i > X_j) = \pi_i/(\pi_i + \pi_j).$$

Let $x_{ij\alpha} = 1$ if $X_i > X_j$ and zero otherwise for the αth comparison of T_i and T_j, $\alpha = 1, \ldots, n_{ij}$, $i, j = 1, \ldots, t$. The likelihood function for the complete experiment, on the assumption of independence between comparisons, is

$$(2.6.1) \qquad L(\pi) = \prod_{i=1}^{t} \prod_{\substack{j=1 \\ j \neq 1}}^{t} \prod_{\alpha=1}^{n_{ij}} \pi_i^{x_{ij\alpha}} \Big/ \prod_{\substack{i,j=1 \\ i<j}}^{t} (\pi_i + \pi_j)^{n_{ij}} .$$

If $n_{ij} = n$ for all i and j, as has been assumed in published work, $L(\pi)$ may be factored

$$(2.6.2) \qquad L(\pi) = \prod_{\alpha=1}^{n} \left[\prod_{i=1}^{t} \prod_{\substack{j=1 \\ j \neq 1}}^{t} \pi_i^{x_{ij\alpha}} \Big/ \prod_{\substack{i,j=1 \\ i<j}}^{t} (\pi_i + \pi_j) \right]$$

and standard ML theorems applied for now $f(x_\alpha)$ may be taken to be the quantity in braces in (2.6.2). If $n_{ij} \neq n$ for some i and j, the methods of this paper for associated populations are required.

Dykstra [26], in work as yet unpublished, has considered extensions of the method for paired comparisons to cases where $n_{ij} \neq n$, and it appears that these are the cases required most frequently in practice in consumer acceptance studies.

III. A SEQUENTIAL DECISION PROCEDURE FOR COMPARING SURVIVAL CURVES

3.1 Introduction. In the search for compounds which may have a useful biological effect, sometimes hundreds of chemicals may be put forth as possible candidates. In this case a preliminary test procedure is needed to select the more promising compounds without resort to long and expensive experimentation. Sequential experimentation immediately suggests itself because of its property of minimizing the average number of observations needed to reach a decision. (See Wald and Wolfowitz [43].)

The need for such a screening procedure arose at Oak Ridge National Laboratory in connection with a search for compounds which might furnish protection against injury caused by irradiation. Preliminary decisions were to be based on the results of treating mice which had been exposed to X-radiation with the experimental compounds. The experimentation was done in groups; five mice being treated with the experimental compound and five with a control compound, which was either a physiological saline or a compound of established protective ability. At first the data collected were the numbers of mice surviving in the control and experimental groups at the end of a prescribed number of days. The period ranged from four to twelve days after irradiation depending on the strain of mice, the radiation dose, and other factors. This procedure proved unsatisfactory since, due to the high variability of the life spans of the irradiated mice, it often happened that in the two groups the animals would all be alive or would all

be dead at the end of the prescribed time period. Obviously results of this nature give no information about the relative merits of the two compounds. To avoid this difficulty, it was decided to stop the experiment at the death of the median (third) mouse in the control groups. Thus no pair of groups of mice are wasted, and all pairs give some information. Of course, a large number of mice surviving in the experimental group tends to indicate the experimental compound is an improvement over the control, and a small number tends to indicate the contrary. Groups were treated successively until the experimental compound was deemed acceptable or not by a statistical criterion. The acceptable compounds were then subjected to more exhaustive and exacting experimentation. The rejection of a compound was not absolute nor necessarily permanent. If a compound chemically related to a rejected compound proved effective, the rejected compound was also subjected to the more exhaustive experimentation.

This sampling procedure was devised and used by Kimball, Burnett, and Doherty [38]. Their method of analyzing the data is the sequential test of the double dichotomy due to Wald [42], p. 106 ff. This method of analysis assumes the probability that a mouse survives in the experimental group at the time the experiment is stopped is constant over the successive trials. This assumption is not valid since the time, t, of stopping the experiment is a random variable; and the probability that a mouse survives depends on the survival curve which in turn is a function of t. Thus, this probability is itself a random variable and the analysis is not theoretically sound as is pointed out in [38].

Although this method of analysis is not rigorously valid, it did prove successful in the search for protective compounds. Thus it was thought to be approximately correct.

Our purpose is to derive rigorously a method of analyzing data collected by using this sampling procedure. It will involve assuming specified functional forms of the survival curves of the control and experimental compounds. We shall limit ourselves to sigmoidal curves, although the methods set forth can be applied to non-sigmoidal curves. The new results will be compared to the original method of analysis; and the OC (Operating Characteristic) function, the ASN (Average Sample Number) function, and the effect of truncation will be considered.

Although the terminology used will involve specific reference to "mice surviving," this may be interpreted as referring to any sort of experimental unit enduring in a prescribed condition.

3.2 Review of Literature. Very little has been written on the application of sequential methods to testing survival curves. Epstein and Sobel [37] have discussed related sampling situations in life testing. However, their situation differs from the one here in several respects. It assumes an exponential survival curve and that the exact lengths of life are recorded. Further, it is not a two sample problem but a test of hypothesis of an exponential parameter. Therefore, their results do not apply.

Two major general sequential procedures have been proposed, those of Rao [41] and Wald [42]. Rao's procedure involves a fixed limit on the sample size and is designed to test null hypotheses. While the former

property is advantageous, the latter is not, since we are not concerned here with testing whether the experimental and control compounds have exactly equal median survival times. Rather, we require a decision procedure for finding experimental compounds with median survival times appreciably better than the control. Further objections to Rao's procedure are that its application requires the derivation of certain properties of the test statistic for the particular case under consideration; and that general methods of deriving the OC function and the ASN function do not exist.

On the other hand, Wald's general procedure is admirably suited to our situation. It is not a test of a null hypothesis, but rather an acceptance-rejection decision procedure which is what we require here. Further, the problem of the derivation of the useful distributional properties of the test statistic has been solved approximately; and general, though approximate, formulas for the OC function, the ASN function, and the effect of truncation are available. Thus we shall apply Wald's procedure to this sampling situation.

3.3 Distribution Theory. Although the distribution of the test statistic is not needed, we must derive certain distributions related to the survival curves before we can apply Wald's general theory. This is due to the peculiarity of our sampling procedure, which, as was stated earlier, makes certain probabilities random variables. In particular, we require the unconditional distribution of the number of mice surviving in the experimental group. We shall proceed to derive this distribution, which will depend, in

in general, on the form of the two survival curves.

Let $p_c(t)$ and $p_\epsilon(t)$ be the survival curves for the control and experimental compounds respectively. Each group consists of 2s + 1 mice, where s is a non-negative integer. Then the distribution of the time of death of the $(\gamma + 1)^{\underline{st}}$ mouse in the control group is[1]

$$(3.3.1) \quad h(t)dt = - \frac{(2s + 1)!}{\gamma!\ (2s - \gamma)!} \left[p_c(t)\right]^{2s-\gamma} \left[1 - p_c(t)\right]^{\gamma} dp_c(t) ,$$

as is shown by, for instance, Cramer [35], p. 368.

Now, as pointed out above, we have chosen to stop the experiment at the death of the median mouse, i.e., $(s + 1)^{\underline{st}}$. Two reasons for this choice may be cited:

(1) We are interested in making decisions based on the comparative sizes of the medians of the two survival curves.

(2) In the cases we shall consider, the sample median time of death is more efficient than any other order statistic. This statement requires some explanation and justification. Cramer [35], p. 369, has shown that the sample quantile, $t_{\gamma+1}$, of order p, where γ is the greatest integer $\leq$ (2s + 1)p, is asymptotically normally distributed with mean equal to T_p, where $p = 1 - p_c(T_p)$, and with variance equal to $(1/-p_c'(T_p))$ $\sqrt{p(1-p)/(2s+1)}$.

[1]The minus sign is required since $p_c(t)$ is equal to one minus the cumulative distribution.

Obviously the coefficient of the radical occurring in the expression for the asymptotic variance is minimized when T_p equals the mode of the distribution, and the radical itself is minimized when $p = 1/2$, that is, when T_p is the population median. Thus, when the underlying distribution is unimodal and symmetric (as, indeed, our final model specifies), the T_p will be both the median and the mode if we choose our order statistic to be the sample median. Since in general using the more efficient statistic leads to the more powerful test, it is thought that this method of sampling yields a test more powerful than one based on any other order statistic.

For the sample median, (3.3.1) becomes

$$(3.3.2) \quad h(t)dt = -\frac{(2s+1)!}{(s!)^2}\left[p_c(t)\right]^s\left[1 - p_c(t)\right]^s dp_c(t) .$$

From this we obtain the distribution of $\Theta = p_\epsilon(t)$, which is

$$(3.3.3) \quad dG(\Theta;\Lambda, s) = \frac{(2s+1)!}{(s!)^2}\left\{p_c\left[p_\epsilon^{-1}(\Theta)\right]\right\}^s\left\{1 - p_c\left[p_\epsilon^{-1}(\Theta)\right]\right\}^s \cdot \frac{p_c'\left[p_\epsilon^{-1}(\Theta)\right]d\Theta}{p_\epsilon'\left[p_\epsilon^{-1}(\Theta)\right]} ,$$

for $U \leq \Theta \leq L$, where $U \geq 0$ and $L \leq 1$. Depending on the ranges of $p_\epsilon(t)$ and $p_c(t)$ either $Pr(\Theta = U)$ or $Pr(\Theta = L)$ may be finite. The Λ appearing in (3.3.3) is a parameter vector whose elements are functions of the parameters

appearing in the two survival curves. This vector will measure in some sense the deviation between the two curves. When $p_c(t) \equiv p_\epsilon(t)$, Λ assumes its null value, Λ^*, and (3.3.3) becomes

$$(3.3.4) \qquad dG(\theta; \Lambda^*, s) = \frac{(2s+1)!}{(s!)^2} \theta^s (1-\theta)^s \, d\theta \, , \; 0 \leq \theta \leq 1,$$

that is, θ has a Beta-distribution. When the group size is one, $s = 0$, and θ has a uniform distribution.

If we let a_r be the number of mice surviving at time t in the $r^{\underline{th}}$ experimental group in a sequence of groups, the conditional distribution of a_r is binomial with parameter θ. The unconditional frequency function is

$$(3.3.5) \qquad f(a_r; \Lambda, s) = \binom{2s+1}{a_r} \int_0^1 \theta^{a_r} (1-\theta)^{2s-a_r+1} \, dG(\theta; \Lambda, s),$$

where $a_r = 0, 1, 2, \ldots, 2s+1$. When $p_\epsilon(t) \equiv p_c(t)$, we have from (3.3.4) that

$$(3.3.6) \qquad f(a_r; \Lambda^*, s) = \binom{2s+1}{a_r} \frac{(2s+1)!}{(s!)^2} \frac{(s+a_r)!\,(3s-a_r+1)!}{(4s+2)!} \, ,$$

$$a_r = 0, 1, \ldots, 2s+1,$$

this result being completely independent of the forms of the survival curves. The frequency function (3.3.6) is symmetric and is more dispersed than the

corresponding binomial with parameter of one-half. This extra-binomial dispersion results from the increased variability introduced by the variate Θ. Of course, when $s = 0$, $f(a_r; \Lambda^*, s)$ reduces to the simple binomial.

When $\Lambda \neq \Lambda^*$, $f(a_r; \Lambda, s)$ will depend on the form of the survival curves, so that we must assume specific forms of $p_c(t)$ and $p_\epsilon(t)$ to obtain the distribution of a_r explicitly.

The reader should note that the derivation of $f(a_r; \Lambda, s)$ outlined above requires that each control group be associated with only one experimental group, since otherwise the probabilities derived are not interpretable in the frequency sense. Further, it can be seen that if several experimental groups are to be compared simultaneously with only one control, any serious departure of the control value from its expectation might lead to a set of simultaneously incorrect conclusions.

3.4 Requirements of the Survival Curve. The results of the previous paragraph are completely general, that is, there is no restriction as to the types of survival curves being considered. We shall now formulate some requirements for survival curves which will make the test practicable from both the statistical and biological points of view.

The forms of $p_\epsilon(t)$ and $p_c(t)$ will be required to satisfy the following conditions:

(1) The vector Λ reduces to a scalar, that is, there is a single parameter λ occurring in the unconditional distribution of a_r which measures the deviation of $p_\epsilon(t)$ from $p_c(t)$. In all the cases which we shall consider, this requirement means that we must assume either that the scale

parameters occurring in $p_\epsilon(t)$ and $p_c(t)$ are equal, or that the location parameters occurring in $p_\epsilon(t)$ and $p_c(t)$ are equal. We shall assume the two scale parameters are equal, since this is the more reasonable biological assumption. Further, this is analogous to the assumption of equal variances in the application of the two sample t test. If one of these assumptions is not made, the application of Wald's sequential analysis requires that the resulting nuisance parameter be integrated out with respect to an optimum weight function. This introduces serious complications into the procedure.

(2) The parameter λ has a meaningful interpretation for the experimenter, since the decision procedure will involve his choosing alternate values for λ. In particular, it is required that λ be monotonic function of $p_\epsilon(m_c)$, where m_c is the median of $p_c(t)$; since the experimenter is usually interested in making decisions based on relative values of the medians.

(3) The forms of $p_\epsilon(t)$ and $p_c(t)$ are theoretically legitimate from the biological point of view. This means that the survival curves should be sigmoidal, and perhaps also, skewed to the right. Further, if $p(t)$ has a range with a lower limit, b, then we require that $p'(b) = 0$.

(4) The decision procedure resulting when requirements (1), (2), and (3) are satisfied leads to tractable results, so that it may be easily applied.

Aside from (3), these requirements are also needed for non-biological applications of the procedure. The principal results presented here shall be based on all four requirements.

3.5 General Statement of the Decision Procedure. Decisions regarding

the acceptance or rejection of a particular experiment compound depend on the possible values of $p_\epsilon(m_c)$. Wald's procedure ([42] p. 78 ff.) requires that the experimenter choose two numbers p_0 and p_1, where $1/2 \leq p_0 \leq 1$ and $0 \leq p_1 \leq 1/2$, such that:

(1) when $p_\epsilon(m_c) \leq p_1$, the experimental compound is appreciably less effective than the control,

(2) when $p_1 \leq p_\epsilon(m_c) \leq p_0$, the difference between the two curves is considered negligible, and

(3) when $p_\epsilon(m_c) \geq p_0$ the experimental compound is considered appreciably better than the control.

These three intervals are termed in order, the zone of preference for rejection, the zone of indifference, and the zone of preference for acceptance. Since λ is required to be a monotonic function of $p_\epsilon(m_c)$, these zones may be defined in terms of λ_0 and λ_1 corresponding to p_0 and p_1 respectively.

Next the experimenter decides on the probabilities of errors which will be tolerated. The probability of rejecting the experimental compound when $p_\epsilon(m_c) = p_0$ is denoted by α, and the probability of accepting the experimental compound when $p_\epsilon(m_c) = p_1$, is denoted by β. The choice of α and β uniquely determines two real numbers, A and B, which are used in the decision procedure.

The test statistic defined by Wald is

$$(3.5.1) \quad \sum_{r=1}^{m} z(a_r) = \sum_{r=1}^{m} \log \frac{f(a_r;\ \lambda_1,\ s)}{f(a_r;\ \lambda_0,\ s)}, \quad a_r = 0, 1, \ldots, 2s + 1;$$

where log denotes a common logarithm. At each stage of the experiment (i.e., at the $m^{\underline{th}}$ trial for each positive integral value of m), the cumulative sum, $\sum_{r=1}^{m} z(a_r)$, is computed. If $\log B < \sum_{r=1}^{m-1} z(a_r) < \log A$ and $\sum_{r=1}^{m} z(a_r) \geq \log A$, the experimental compound is rejected. If $\log B < \sum_{r=1}^{m} z(a_r) < \log A$, an additional trial is performed. If $\log B < \sum_{r=1}^{m-1} z(a_r) < \log A$ and $\sum_{r=1}^{m} z(a_r) \leq \log B$, the experimental compound is accepted. The process of repeating trials could conceivably go on indefinitely, but it has been shown ([42], p. 157 ff.) that the probability is one that the procedure will terminate at some finite m. In spite of this, it may happen that testing a given compound will lead to a large number of trials. Thus, for economic or other reasons, it may be necessary to stop the trials before a decision is reached by the above rule. If we truncate the procedure in this way, say at the $m^{\underline{th}}$ trial, the experimental compound is rejected if $\sum_{r=1}^{m} z(a_r) > 0$ and it is accepted if $\sum_{r=1}^{m} z(a_r) < 0$. This truncation will affect the values of the probabilities of errors involved, making them larger than the stated values. This change will be investigated in paragraph 3.15.

Although A and B are uniquely determined by α and β, simple general expressions for them have not been found. Wald [42], p. 44 ff., recommends approximating A by $(1 - \beta)/\alpha$ and B by $\beta/(1 - \alpha)$. Then the true probabilities of errors of the first and second kind respectively do not exceed their prescribed values, α and β. Further, the use of these values causes a slight increase in the ASN function.

3.6 The $f(a_r; \lambda, s)$ when $p(t) = e^{-k(t-b)^2}$. The first three requirements

for the survival curve cited in paragraph 3.4 are satisfied if we let,

$$p_c(t) = 1, \qquad t < b_c ,$$

$$p_c(t) = e^{-k(t-b_c)^2}, \qquad t \geq b_c ,$$

and

$$p_\epsilon(t) = 1, \qquad t < b_\epsilon ,$$

$$p_\epsilon(t) = e^{-k(t-b_\epsilon)^2}, \qquad t \geq b_c ,$$

where k, b_c, and b_ϵ are all constants greater than zero. The scale parameter k is assumed equal in the two curves in order that only one parameter, $\lambda = \sqrt{k}\,(b_\epsilon - b_c)$, occur in the distribution of Θ. Further,

$$m_c = b_c + \frac{(\ln 2)^{1/2}}{k^{1/2}} ,$$

and

$$p_\epsilon(m_c) = e^{-\left[(\ln 2)^{1/2} - \lambda\right]^2} ,$$

so that

$$\lambda = (\ln 2)^{1/2} - \left\{ - \ln p_\epsilon(m_c) \right\}^{1/2}, \; -\infty < \lambda < (\ln 2)^{1/2} .$$

Obviously λ is a monotonic function of $p_\epsilon(m_c)$, so that λ and $p_\epsilon(m_c)$ are in one to one correspondence.

Though otherwise satisfactory, this form of survival curve leads to

an intractable expression for $f(a_r;\ \lambda,\ s)$. Therefore its derivation will merely be outlined. The derivation is divided into three cases depending on whether $\lambda = \lambda_0 > 0$, $\lambda = \lambda^* = 0$, or $\lambda = \lambda_1 < 0$.

Case 1. $\lambda = \lambda_0 > 0$, $(p_\epsilon(m_c) = p_0 > 1/2)$. Following the derivation indicated in paragraph[2] 3.3,

$$dG(1;\ \lambda_0,\ s) = I_{1-\exp(-\lambda^2)}(s+1,\ s+1),$$

and

$$dG(\theta;\ \lambda_0,\ s) = \frac{(2s+1)!}{(s!)^2}\left\{1 - e^{-\left[(1-\ln\theta)^{1/2}+\lambda_0\right]^2}\right\}^s \cdot$$

$$\left\{e^{-\left[(-\ln\theta)^{1/2}+\lambda_0\right]^2}\right\}^{s+1}\left\{1+\frac{\lambda_0}{(-\ln\theta)^{1/2}}\right\}\frac{d\theta}{\theta},\quad 0 \leq \theta < 1.$$

So that in order to derive $f(a_r;\ \lambda_0,\ s)$, we must evaluate the integral:

$$\int_0^1 \theta^{a_r}(1-\theta)^{2s-a_r+1}\left\{1 - e^{-\left[(-\ln\theta)^{1/2}+\lambda_0\right]^2}\right\}^s$$

$$\left\{e^{-\left[(-\ln\theta)^{1/2}+\lambda_0\right]^2}\right\}^{s+1}\left\{1+\frac{\lambda_0}{(-\ln\theta)^{1/2}}\right\}\frac{d\theta}{\theta}.$$

---

[2]Following Pearson [40], vi, we let

$$I_x(p,\ q) = \left[B(p,\ q)\right]^{-1}\int_0^x t^{p-1}(1-t)^{q-1}\,dt.$$

To evaluate this integral, we make the change of variable $x = (-\ln \theta)^{1/2}$ and expand the two binomial expressions occurring in binomial series. Completing the square on the quadratic exponents puts the integrands in the form of normal densities, so that we have finally,[3]

$$f(a_r;\ \lambda_0,\ s) = \frac{(2s+1)!}{(s!)^2} \binom{2s+1}{a_r} \sum_{i=0}^{2s-a_r+1} \sum_{j=0}^{s} \left\{ \binom{2s-a_r+1}{i} \binom{s}{j} (-1)^{i+j} \right.$$

$$\left. \cdot \frac{e^{-\lambda_0^2 D_j (C_{ij}-D_j)/C_{ij}}}{C_{ij}} \left[e^{-\lambda_0^2 D_j^2/C_{ij}} + \frac{2\sqrt{\pi}\,\lambda_0 (C_{ij}-D_j)}{C_{ij}^{1/2}} \Phi\left(\frac{-2\lambda_0 D_j}{C_{ij}^{1/2}} \right) \right] \right\},$$

$$a_r = 0,\ 1,\ \ldots,\ 2s,$$

where

$$C_{ij} = s + a_r + i + j + 1, \text{ and } D_j = s + j + 1.$$

Further

$$f(2s+1;\ \lambda_0,\ s) = I_{1-e^{-\lambda_0^2}}(s+1,\ s+1) + \frac{(2s+1)!}{(s!)^2} \sum_{j=0}^{s} \left\{ \binom{s}{j} (-1)^j \right.$$

---

[3]We let

$$\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

$$\cdot \frac{e^{-\lambda_0 D_j (C_j' - D_j)/C_j'}}{C_j'} \left[e^{-\lambda_0^2 D_j^2 / C_j'} + \frac{2\sqrt{\pi}\,(C_j' - D_j)}{(C_j')^{1/2}} \Phi\left(\frac{\lambda_0\, 2D_j}{(C_j')^{1/2}} \right) \right] \Bigg\},$$

where

$$C_j' = 3s + j + 2.$$

Case 2. $\lambda = \lambda^* = 0$, $(p_\epsilon(m_c) = 1/2)$. This implies $p_c(t) = p_\epsilon(t)$ so that $dG(\theta); \lambda^*, s)$ is the Beta-distribution and $f(a_r; \lambda^*, s)$ is given by (3.3.6).

Case 3. $\lambda = \lambda_1 < 0$, $(p_\epsilon(m_c) = p_1 < 1/2)$. Similar to case 1, we have

$$dG(\theta; \lambda_1, s) = 0, \; 1 \geq \theta > e^{-\lambda_1^2}$$

$$dG(\theta; \lambda_1, s) = \frac{(2s+1)!}{(s!)^2} \left\{ 1 - e^{-\left[(-\ln\theta)^{1/2} + \lambda_1\right]^2} \right\}^s$$

$$\cdot \left\{ e^{-\left[(-\ln\theta)^{1/2} + \lambda_1\right]^2} \right\}^{s+1} \left\{ 1 + \frac{1}{(-\ln\theta)^{1/2}} \right\} \frac{d\theta}{\theta} \quad ,$$

$$0 \leq \theta \leq e^{-\lambda_1^2} .$$

Proceeding as in case 1, we obtain

$$f(a_r;\lambda_1, s) = \frac{(2s+1)!}{(s!)^2} \quad \frac{(2s+1)!}{(s!)^2} \sum_{i=0}^{2s-a_r+1} \sum_{j=0}^{s} \left\{ \binom{2s-a_r+1}{i} \binom{s}{j} (-1)^{i+j} \right.$$

$$\left. \cdot \frac{e^{-\lambda_0^2 D_j (C_{ij} - D_j)/C_{ij}}}{C_{ij}} \left[e^{-\lambda_1^2 (D_j+1)^2/C_{ij}} + \frac{2\sqrt{\pi}\lambda_1 (C_{ij} - D_j)}{C_{ij}^{1/2}} \Phi\left(\frac{-\sqrt{2}\lambda_1 (D_j+1)}{C_{ij}^{1/2}} \right) \right] \right\},$$

$$a_r = 0, 1, \ldots, 2s+1.$$

Easy application of Wald's procedure requires specific numerical results for $f(a_r; \lambda, s)$ for various values of λ. But as can be seen, explicit results are quite difficult to compute except when $s = 0$. Therefore, this model was discarded.

Assuming $p(t) = \operatorname{sech}^2 k(t - b)$, $t > b$, also leads to intractable results for the distribution of $f(a_r; \lambda, s)$. Therefore, we turn to a model which does not satisfy exactly all the requirements cited above, but does lead to tractable results for small group sizes.

3.7 Discussion of the Logistic Curve. We consider now the density

$$p(t) = \frac{\pi}{4\sqrt{3}\sigma} \operatorname{sech}^2 \frac{\pi}{\sqrt{3}} \left(\frac{t-\mu}{\sigma}\right) \qquad -\infty < t < \infty ,$$

where $\sigma > 0$ and μ may be positive or negative. The graph of the cumulative distribution of this density, which is called the logistic curve, has been extensively investigated by Winsor [44]. He has shown that the mean and variance are μ and σ^2 respectively, and that

$$E(t-\mu)^{2n+1} = 0, \quad n = 0, 1, 2, \ldots,$$

and

$$E(t-\mu)^{2n} = 3^n(2^{2n} - 2)\ \sigma^{2n}\ B_n, \quad n = 1, 2, 3, \ldots,$$

where B_n is the Bernoulli number of order n. When $n = 2$, $\mu_4 = 21\sigma^4/5$, and the kurtosis is

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6}{5} \ :$$

Since $\gamma_2 > 0$, the logistic density is leptokurtic, that is, sharply peaked relative to the normal density. Like the normal density the logistic density is symmetric and Winsor has shown that the logistic curve is practically indistinguishable from the integrated normal curve the same mean and variance.

If we assume the survival times of the mice have a logistic density, the corresponding survival curve is

$$p(t) = 1/2 \left[1 - \tanh \frac{\pi}{\sqrt{3}} \left(\frac{t-\mu}{\sigma}\right) \right] = \frac{1}{1+e^{\pi(t-\mu)/\sqrt{3}\sigma}}, \ -\infty < t < \infty \ .$$

An immediate objection to the model is that it assigns finite probabilities to negative survival times. If we attempt to circumvent this difficulty by truncating the distribution at $t = 0$ and adjusting the multiplicative constant accordingly, the Λ which is involved in $dG(\theta;\Lambda, s)$ does not reduce to a scalar. Therefore the distribution was left untruncated. This

objection is not too serious for two reasons.

(1) The probability that t is less than zero is quite small when μ/σ is large. Specifically, we have

$$\Pr(t < 0;\ \mu/\sigma) = \frac{1}{1+e^{\pi\mu/\sqrt{3}\sigma}} ,$$

so that when $\mu/\sigma = 2$, $\Pr(t < 0) = 0.027$; and when $\mu/\sigma = 5$, $\Pr(t < 0) = 0.0001$. Thus this probability goes to zero quite rapidly as μ/σ becomes large.

(2) The procedure based on this model is relatively insensitive to the form of sigmoidal curve assumed. This property will be investigated extensively in paragraph 3.11.

On the other hand, this form of survival curve does possess many desirable properties. It has the property that $\lim_{t \to \infty} \left\{ p'(t) \right\} = 0$. Further, it leads to a general expression for $f(a_r;\ \lambda,\ s)$ which can be relatively easily tabled for $s = 0, 1, 2$. The distribution of a_r also possesses certain symmetry properties which aid in the computations involved in finding the OC and ASN functions. Therefore, we shall investigate this model extensively.

3.8 Derivation of $f(a_r;\ u,\ s)$ for the Logistic Model. Let

$$p_\epsilon(t) = \frac{1}{1+e^{\pi(t-\mu_\epsilon)/\sqrt{3}\sigma_\epsilon}} ,\quad -\infty < t < \infty ,$$

and

$$p_c(t) = \frac{1}{1+e^{\pi(t-\mu_c)/\sqrt{3}\sigma_c}}, \quad -\infty < t < \infty .$$

Letting $\theta = p_\epsilon(t)$, we have,

$$t = \mu_\epsilon + \frac{\sqrt{3}\sigma_\epsilon}{2\pi} \ln\left(\frac{1-\theta}{\theta}\right)$$

which, upon making the change of variable in the distribution of the median time of death, i.e., (3.3.1), gives the result corresponding to (3.3.3); namely,

$$(3.8.1) \quad dG(\theta;\Lambda, s) = \frac{(2s+1)!}{(s!)^2} \lambda_1 \frac{\left[e^{(2\pi\lambda_2/\sqrt{3})(s+1)}\right]\left(\frac{1-\theta}{\theta}\right)^{\lambda_1(s+1)}}{\left[e^{2\pi\lambda_2/\sqrt{3}}\left(\frac{1-\theta}{\theta}\right)^{\lambda_1} + 1\right]^{2s+2}} \frac{d\theta}{\theta(1-\theta)},$$

$$0 \leq \theta \leq 1,$$

where $\Lambda = [\lambda_1, \lambda_2] = [\sigma_\epsilon/\sigma_c, (\mu_\epsilon-\mu_c)/\sigma_c]$. If either $\mu_c = \mu_\epsilon$ or $\sigma_c = \sigma_\epsilon$, the vector Λ reduces to a scalar which is a monotonic function of $p_\epsilon(\mu_c)$. As was pointed out in paragraph 3.4, it is more reasonable to assume $\sigma_\epsilon = \sigma_c$, and this leads to

$$dG(\theta; u, s) = \frac{(2s+1)!}{(s!)^2} \frac{u^{s+1}}{(1-u)^{2s+2}} \frac{\theta^s(1-\theta)^2}{\left[\frac{u}{1-u} + \theta\right]^{2s+2}} d\theta, \quad 0 \leq \theta \leq 1,$$

where $u = e^{2\pi\lambda_2/\sqrt{3}} = p_\epsilon(\mu_c)/\left[1 - p_\epsilon(\mu_c)\right]$. In the following discussion we shall write the distribution of a_r as a function of either u or p, where $p = p_\epsilon(\mu_c)$.

The distribution of a_r for the logistic model, which we denote by $f_L(a_r; u, s)$ is found from (3.3.5) to be,

$$(3.8.2) \quad f_L(a_r; u, s) = \frac{(2s+1)!}{(s!)^2} \binom{2s+1}{a_r} \frac{1}{u^{s+1}} \int_0^1 \frac{\theta^{s+a_r}(1-\theta)^{3s-a_r+1}}{\left[1 - (\frac{u-1}{u})\theta\right]^{2s+2}} \, d\theta ,$$

$$a_r = 0, 1, \ldots, 2s+1.$$

The integral in (3.8.2) is functionally related to the integral form of the hypergeometric function, and MacRobert [39], p. 297, shows that the integral is equal to[4]

$$(3.8.3) \quad B(s+a_r+1, 3s-a_r+2) \; {}_2F_1 \left[2s+2, s+a_r+1, 4s+3, (u-1)/u \right]$$

when $|(u-1)/u| < 1$, which implies $u > 1/2$.

Since ${}_2F_1 \left[\alpha, \beta, \gamma, z\right] = (1-z)^{-\alpha} \; {}_2F_1 \left[\alpha, \gamma-\beta, \gamma, z/(z-1)\right]$, (see [39], p. 297), we may write for (3.8.3)

$$(3.8.4) \quad B(s+a_r+1, 3s-a_r+2) \, u^{2s+2} \; {}_2F_1 \left[2s+2, 3s-a_r+2, 4s+3, 1 - u \right].$$

---

[4] Some authors omit the subscripts from before and after F when denoting the hypergeometric function.

Substituting this result into (3.8.2), we have

$$(3.8.5)\quad f_L(a_r;\, u,\, s) = \frac{(2s+1)!}{(s!)^2} \binom{2s+1}{a_r} B(s+a_r+1,\, 3s+a_r+2)\, u^{s+1}$$

$$\cdot\, {}_2F_1\left[\, 2s+2,\, 3s-a_r+2,\, 4s+3,\, 1-u \,\right],\; u > 1/2,\; a_r = 0,\, 1,\, \ldots\, 2s+1.$$

Unfortunately the hypergeometric function has not been tabled for the arguments in which we are interested, so this expression is not immediately useful.

Let us return to (3.8.2) and make the transformation $x = \theta/(1-\theta)$ in the integral appearing there. This gives

$$(3.8.6)\quad f_L(a_r;\, u,\, s) = \frac{(2s+1)!}{(s!)^2} \binom{2s+1}{a_r} u^{s+1} \int_0^\infty \frac{x^{s+a_r}\, dx}{(x+1)^{2s+1}(x+u)^{2s+2}} .$$

From this form of the integral it can easily be shown that

$$(3.8.7)\qquad f_L(a_r;\, u,\, s) = f_L(2s+a_r+1;\, u^{-1},\, s),$$

by letting $x = 1/y$ in the integral occurring on the right. This relation permits us to use (3.8.5) for finding $f_L(a_r;\, u,\, s)$ when $0 < u \leq 1/2$. Since $u = p/(1-p)$, (3.8.7) implies that

$$(3.8.8)\qquad f_L(a_r;\, p,\, s) = f_L(2s+a_r+1;\, 1-p,\, s) .$$

This relation reduces considerably the computations involved in tabling the distribution of a_r.

Returning to (3.8.6), we see that the integrand is a rational function whose denominator has only negative zeros and is of degree $4s+3$, while the numerator is at most of degree $3s+1$. Therefore, we may apply the theorem of appendix B, and write the integral in (3.8.6) as

$$(3.8.9)\quad -\frac{1}{2\pi i}\int_{C_{-1}} \frac{z^{s+a_r}(z+u)^{-(2s+2)}\ln z\,dz}{(z+1)^{2s+1}} - \frac{1}{2\pi i}\int_{C_{-u}} \frac{z^{s+a_r}(z+1)^{-(2s+1)}\ln z\,dz}{(z+u)^{2s+2}}$$

where C_{-1} and C_{-u} are small disjoint circles with centers at -1 and $-u$, and the integrations are performed in a positive direction with respect to the centers.

Making use of the integral representation of the derivative of a function (see [39], p. 70), we may write (3.8.9) in the form,

$$(3.8.10)\quad -\frac{1}{(2s)!}\frac{d^{2s}}{dz^{2s}}\left\{ z^{s+a_r}(z+u)^{-(2s+2)}\ln z \right\}\Big|_{z=-1}$$

$$-\frac{1}{(2s+1)!}\frac{d^{2s+1}}{dz^{2s+1}}\left\{ z^{s+a_r}(z+1)^{-(2s+1)}\ln z \right\}\Big|_{z=-u} .$$

By Leibniz's formula, this equals,

$$(3.8.11)\quad - \sum_{j+k+m=2s} \left\{ \frac{1}{j!k!m!} \frac{d^j z^{s+a_r}}{dz^j} \frac{d^k}{dz^k} \ln z \frac{d^m}{dz^m} (z+u)^{-(2s+2)} \right\} \Bigg|_{z=-1}$$

$$- \sum_{j+k+m=2s+1} \left\{ \frac{1}{j!k!m!} \frac{d^j z^{s+a_r}}{dz^j} \frac{d^k \ln z}{dz^k} \frac{d^m}{dz^m} (z+1)^{-(2s+1)} \right\} \Bigg|_{z=-u},$$

where the summations are taken over all the non-negative integers, j, k, and m, subject to the restrictions noted under the summation symbols.

To evaluate (3.8.11) we need the following expressions,

$$(3.8.12)\quad \begin{cases} \dfrac{d^j z^{s+a_r}}{dz^j} \Bigg|_{z=-1} = \dfrac{(s+a_r)!}{(s+a_r-j)!} (-1)^{s+a_r-j}, \; j \geq 0 \\[2ex] \dfrac{d^j z^{s+a_r}}{dz^j} \Bigg|_{z=-u} = \dfrac{(s+a_r)!}{(s+a_r-j)!} (-1)^{s+a_r-j} u^{s+a_r-j}, \; j \geq 0, \end{cases}$$

$$(3.8.13)\quad \begin{cases} \left.\dfrac{d^k \ln z}{dz^k}\right|_{z=-1} = -(k-1)! \quad , \; k > 0, \\[2ex] \left.\dfrac{d^k \ln z}{dz^k}\right|_{z=-u} = -(k-1)!\, u^{-k} \; , \quad k > 0, \\[2ex] \left.\dfrac{d^o \ln z}{dz^o}\right|_{z=-1} = \pi i \quad , \\[2ex] \left.\dfrac{d^o \ln z}{dz^o}\right|_{z=-u} = \ln u + \pi i \end{cases}$$

$$(3.8.14)\quad \begin{cases} \left.\dfrac{d^m}{dz^m}\,(z+u)^{-(2s+2)}\right|_{z=-1} = \dfrac{(2s+m+1)!}{(2s+1)!\,(1-u)^{2s+m+2}} \; , \; m \geq 0, \\[2ex] \left.\dfrac{d^m}{dz^m}\,(z+1)^{-(2s+1)}\right|_{z=-u} = \dfrac{(-1)^m\,(2s+m)!}{(2s)!\,(1-u)^{2s+m+1}} \quad m \geq 0. \end{cases}$$

Consider now the terms of (3.8.11) which are generated when $k = 0$ in the two triple sums. Using the expression just derived, we may write these terms in the form of two simple summations,

$$(3.8.15)\quad -(s+a_r)!\left\{\pi i \sum_{m=0}^{2s} \frac{(-1)^{a_r+m-s}}{(2s-m)!\,m!}\,\frac{(2s+m+1)!}{(a_r+m-s)!\,(2s+1)!}\,\frac{1}{(1-u)^{2s+m+2}}\right.$$

$$\left. + (\pi i + \ln u) \sum_{m=0}^{2s+1} \frac{(-1)^{a_r+m-s-1}}{(2s-m+1)!\,m!}\,\frac{(2s+m)!}{(a_r+m-s-1)!\,(2s)!}\,\frac{1}{(1-u)^{2s+m+1}}\right\}.$$

Since the integral is a real number and all other terms in (3.8.11) are real, the coefficient of i must be zero, whereupon the negative of the first summation must equal the second summation. Thus the first summation may be used as a factor of the coefficient of $\ln u$.

Using this result together with the results of (3.8.12) - (3.8.14) in (3.8.11), we find from (3.8.6), that

$$(3.8.16)\quad f_L(a_r;\, u,\, s) = \binom{2s+1}{a_r}\frac{(s+a_r)!}{(s!)^2}\cdot\frac{u^{s+1}}{(1-u)^{2s+1}}$$

$$\cdot\left\{\frac{1}{(1-u)}\sum_{\substack{j+k+m=2s\\ k\neq 0}}\left[\frac{(-1)^{s+a_r+j}}{j!\,m!\,k}\,\frac{(2s+m+1)!}{(s+a_r-j)!}\,\frac{1}{(1-u)^m}\right] +\right.$$

$$(2s+1)\,u^{s+a_r}\sum_{\substack{j+k+m=2s+1\\ k\neq 0}}\left[\frac{(-1)^{s+a_r+j+m}}{j!\,m!\,k}\cdot\frac{(2s+m)!}{(s+a_r-j)!}\right.\cdot$$

$$\left.\frac{1}{(1-u)^m\, u^{j+k}}\right] + \frac{1}{(1-u)}\left[\sum_{m=0}^{2s} \frac{(-1)^{s+a_r+m}}{(2s-m)!\,m!}\;\frac{(2s+m+1)!}{(a_r+m-s)!}\cdot\frac{1}{(1-u)^m}\right]\ln u\;\Bigg\}\,,$$

$$u > 0, \text{ and } a_r = 0, 1, \ldots 2s + 1.$$

This expression has been used to derive explicit formulas for $f_L(a_r; u, s)$, for $s = 0, 1, 2$. The results appear in appendix C. These formulas in turn were used to derive $f_L(a_r; p, s)$ for p at intervals of one-tenth. The latter numerical results appear in Table 1, where the tabling makes use of the symmetry relation (3.8.8). When $p = 1/2$, $f_L(a_r; p, s)$ reduces to (3.3.6), and the numerical result is given in Table 2.

For $s > 2$, (3.8.16) becomes quite cumbersome since for each value of a_r, the $f_L(a_r; p, s)$ contains $2(2s + 1)(s + 1)$ terms (a few of which vanish). Thus for $s = 3$, the number of terms for each a_r is thirty; and since the range of a_r then has eight distinct values, two hundred and forty terms must be evaluated to determine $f_L(a_r; p, 3)$, $a_r = 0, 1, 2, \ldots 7$. Thus for larger values of s, it may be simpler to evaluate $f_L(a_r; p, s)$ by using (3.8.5) with the series expansion of the hypergeometric function on a high speed computer. Of course, if the hypergeometric function were to be tabled for the appropriate arguments, tables of $f_L(a_r; p, s)$ could be computed directly from them using (3.8.5)

3.9 A Linear Approximation to the Survival Curve. Though we have results which are easily applied, they are strictly appropriate only when the underlying distributions are of the logistic form. It would be of some

Table 1

TABLE OF $f_L(a_r; p, s)$ FOR $s = 0, 1, 2$.

$s = 0$						
p	0	0.1	0.2	0.3	0.4	
$f_L(0; p, 0)$	1.0000	0.8160	0.7172	0.6379	0.5672	$f_L(1; 1-p, 0)$
$f_L(1; p, 0)$	0.0000	0.1840	0.2828	0.3621	0.4328	$f_L(0; 1-p, 0)$
$s = 1$						
$f_L(0; p, 1)$	1.0000	0.6729	0.4958	0.3714	0.2761	$f_L(3; 1-p, 1)$
$f_L(1; p, 1)$	0.0000	0.2441	0.3162	0.3355	0.3269	$f_L(2; 1-p, 1)$
$f_L(2; p, 1)$	0.0000	0.0682	0.1425	0.2065	0.2592	$f_L(1; 1-p, 1)$
$f_L(3; p, 1)$	0.0000	0.0148	0.0455	0.0866	0.1378	$f_L(0; 1-p, 1)$
$s = 2$						
$f_L(0; p, 2)$	1.0000	0.5610	0.3498	0.2222	0.1391	$f_L(5; 1-p, 2)$
$f_L(1; p, 2)$	0.0000	0.2936	0.3241	0.2911	0.2373	$f_L(4; 1-p, 2)$
$f_L(2; p, 2)$	0.0000	0.1050	0.1942	0.2405	0.2528	$f_L(3; 1-p, 2)$
$f_L(3; p, 2)$	0.0000	0.0314	0.0908	0.1512	0.2026	$f_L(2; 1-p, 2)$
$f_L(4; p, 2)$	0.0000	0.0077	0.0332	0.0730	0.1222	$f_L(1; 1-p, 2)$
$f_L(5; p, 2)$	0.0000	0.0013	0.0079	0.0220	0.0460	$f_L(0; 1-p, 2)$

interest to know the extent to which the probabilities of errors of the first and second kind are altered when the results just obtained are applied to populations having distributions other than the logistic. In other words it would seem advisable to derive the distribution of a_r for a sigmoidal curve satisfying the requirements of paragraph 3.4 and then to compare the results with those for the logistic curve. Unfortunately, as was seen in 3.6, all attempts at the derivation for sigmoidal curves other than the logistic led to intractable results. As an alternative, we turn to a linear approximation to the survival curve.

Any survival curve, $p(t)$, possessing derivatives of all orders at its median, m, may be expanded in a Taylor's series

$$p(t) = \sum_{i=0}^{\infty} p^{(i)}(m) \frac{(t - m)^i}{i!} .$$

If we neglect all but the linear term, we have an approximation to the survival curve, $p^*(t)$, where

$$p^*(t) = 1, \quad t < m + \frac{1}{2p'(m)} ,$$

$$p^*(t) = 1/2 + p'(m)\,(t - m), \quad m + \frac{1}{2p'(m)} \leq t \leq m - \frac{1}{2p'(m)} ,$$

$$p^*(t) = 0, \quad t > m - \frac{1}{2p'(m)} .$$

The approximation has the same median and the same slope at the

median as $p(t)$. However, it has a finite range, since the underlying density is the uniform. We proceed to investigate this approximation.

3.10 Derivation of $f(a_r; p, s)$ for the Linear Model. Define $p_c(t)$ and $p_\epsilon(t)$, such that

$$p_c^*(t) = 1, \qquad t < m_c - \frac{k}{2}$$

$$p_c^*(t) = 1/2 - k(t - m_c), \quad m_c - \frac{k}{2} \leq t \leq m_c + \frac{k}{2}$$

$$p_c^*(t) = 0, \qquad t > m_c + \frac{k}{2}$$

and

$$p_\epsilon^*(t) = 1, \qquad t < m_\epsilon - \frac{k}{2}$$

$$p_\epsilon^*(t) = 1/2 - k(t - m_\epsilon), \quad m_\epsilon - \frac{k}{2} \leq t \leq m_\epsilon + \frac{k}{2}$$

$$p_\epsilon^*(t) = 0, \qquad t > m_\epsilon + \frac{k}{2} \ .$$

We have let $p_c'(m_c) = p_\epsilon'(m_c) = -k$, $k > 0$, since this is equivalent to letting the scale parameters be equal in the logistic case.

Since m_c is the median,

$$p_\epsilon^*(m_c) = 1, \qquad m_c < m_\epsilon - \frac{k}{2}$$

$$p_\epsilon^*(m_c) = 1/2 - k(m_c - m_\epsilon), \quad m_\epsilon - \frac{k}{2} \leq m_c \leq m_\epsilon + \frac{k}{2}$$

$$p_\epsilon^*(m_c) = 0, \qquad m_c > m_\epsilon + \frac{k}{2} \ .$$

We let $\lambda = k(m_\epsilon - m_c)$, $-1/2 < \lambda < 1/2$, so that $\lambda = p_\epsilon(m_c) - 1/2$. Three cases must be considered, $\lambda = \lambda_0 > 0$, $\lambda = \lambda^* = 0$, and $\lambda = \lambda_1 < 0$.

Case 1. $\lambda = \lambda_0 > 0$ $(p_\epsilon(m_c) = p_0 > 1/2)$. The probability that $\Theta = 1$ is equal to the probability that t in (3.3.2) is less than $m_\epsilon - k/2$. Integrating (3.3.2) from $m_c - k/2$ to $m_\epsilon - k/2$, and making the change of variable $\Theta = p_c(t)$, we have

$$dG(1;\ \lambda_0,\ s) = I_{\lambda_0}(s + 1,\ s + 1).$$

As in (3.3.3), we have

$$dG(\Theta;\ \lambda_0,\ s) = \frac{(2s + 1)!}{(s!)^2}\ (\Theta - \lambda_0)^s\ (1 - \Theta + \lambda_0)^s\ d\Theta,\ \lambda_0 \leq \Theta < 1,$$

and

$$dG(\Theta;\ \lambda_0,\ s) = 0, \quad \Theta < \lambda_0.$$

Indicating the distribution of a_r when a uniform density is assumed by $f_U(a_r;\ \lambda_0,\ s)$, we have

$$f_U(a_r;\ \lambda_0,\ s) = \binom{2s+1}{a_r}\int_0^1 \Theta^{a_r}(1 - \Theta)^{2s-a_r+1}\ dG(\Theta;\ \lambda_0,\ s).$$

For $a_r < 2s + 1$, this becomes

$$f_U(a_r;\ \lambda_0,\ s) = \frac{(2s+1)!}{(s!)^2}\binom{2s+1}{a_r}\int_0^1 \theta^{a_r}(1-\theta)^{2s-a_r+1}(\theta-\lambda_0)^s\left[(1-\theta)+\lambda_0\right]^2 d\theta.$$

If we expand the binomial expressions containing in binomial series, the integral may be expressed as a sum of incomplete Beta functions, and we have

$$(3.10.1)\quad f_U(a_r;\ s,\ \lambda_0) = \frac{(2s+1)!}{(s!)^2}\binom{2s+1}{a_r}\sum_{i=0}^{s}\sum_{j=0}^{s}\left\{\binom{s}{j}\binom{s}{j}(-1)^i\ \lambda_0^{i+j}\right.$$

$$\left.\cdot\ B(3s-a_r-j+2,\ s+a_r-i+1)\ \ I_{1-\lambda_0}(3s-a_r-j+2,\ s+a_r-i+1)\right\},$$

$$a_r = 0,\ 1,\ \ldots,\ 2s.$$

Similarly, for $a_r = 2s+1$, we have

$$(3.10.2)\quad f_U(2s+1;\ s,\ \lambda_0) = I_{\lambda_0}(s+1,\ s+1) + \frac{(2s+1)!}{(s!)^2}\sum_{i=0}^{s}\sum_{j=0}^{s}\left\{\binom{s}{j}\binom{s}{j}(-1)^i\ \lambda_0^{i+j}\right.$$

$$\left.\cdot\ B(s-j+1,\ 3s-i+2)\ \ I_{1-\lambda_0}(s-j+1,\ 3s-i+2)\right\}.$$

Case 2. $\lambda = \lambda^* = 0$ $(p_\epsilon(m_c) = p = 1/2)$. This implies $p_c(t) = p_\epsilon(t)$ so that $dG(\theta;\ \lambda^*,\ s)$ is the Beta-distribution and $f(a_r;\ \lambda^*,\ s)$ is given by (3.3.6).

Case 3. $\lambda = \lambda_1 < 0(p_\epsilon(m_c) = p_1 < 1/2)$. Proceeding as in case 1, we have,

$$dG(\theta;\ \lambda_1,\ s) = 0, \quad 1 + \lambda_1 \leq \theta \leq 1,$$

$$dG(\theta;\ \lambda_1,\ s) = \frac{(2s+1)!}{(s!)^2}\,(\theta - \lambda_1)^s(1 - \theta + \lambda_1)^s\, d\theta, \quad 0 < \theta \leq 1 + \lambda_1,$$

and

$$dG(0;\ \lambda_1,\ s) = I_{-\lambda_1}(s + 1,\ s + 1).$$

Further,

$$(3.10.3) \quad f_U(0;\ \lambda_1,\ s) = I_{-\lambda_1}(s+1,\ s+1) + \frac{(2s+1)!}{(s!)^2} \sum_{i=0}^{s} \sum_{j=0}^{s} \binom{s}{i}\binom{s}{j} (-1)^j$$

$$\cdot\ \lambda_1^{i+j}\ B(s-j+1,\ 3s+i+2)\ I_{1+\lambda_1}(s+j+1,\ 3s-i+2),$$

and

$$(3.10.4) \quad f_U(a_r;\ \lambda_1,\ s) = \frac{(2s+1)!}{(s!)^2}\binom{2s+1}{a_r} \sum_{i=0}^{s} \sum_{j=0}^{s} \left\{ \binom{s}{i}\binom{s}{j} (-1)^j\ \lambda_1^{i+j} \right.$$

$$\left. \cdot\ B(s+a_r-j+1,\ 3s-a_r-i+2)\ \ I_{1+\lambda_1}(s+a_r-j+1,\ 3s-a_r-i+2) \right\},$$

$$a_r = 1,\ 2,\ \ldots,\ 2s+1.$$

By comparing (3.10.1) with (3.10.4) and (3.10.2) with (3.10.3), we can easily show that

Table 2

TABLE OF $f_U(a_r; p, s)$ FOR $s = 0, 1, 2$

$s = 0$					
p	0.2	0.3	0.4	0.5	
$f_U(0; p, 0)$	0.7550	0.6800	0.5950	0.5000	$f_U(1; 1-p, 0)$
$f_U(1; p, 0)$	0.2450	0.3200	0.4050	0.5000	$f_U(0; 1-p, 0)$
$s = 1$					
$f_U(0; p, 1)$	0.5518	0.4241	0.3030	0.2000	$f_U(3; 1-p, 1)$
$f_U(1; p, 1)$	0.2662	0.3023	0.3151	0.3000	$f_U(2; 1-p, 1)$
$f_U(2; p, 1)$	0.1433	0.2015	0.2578	0.3000	$f_U(1; 1-p, 1)$
$f_U(3; p, 1)$	0.0387	0.0721	0.1240	0.2000	$f_U(0; 1-p, 1)$
$s = 2$					
$f_U(0; p, 2)$	0.4333	0.2826	0.1626	0.0833	$f_U(5; 1-p, 2)$
$f_U(1; p, 2)$	0.2495	0.2570	0.2307	0.1786	$f_U(4; 1-p, 2)$
$f_U(2; p, 2)$	0.1796	0.2249	0.2467	0.2381	$f_U(3; 1-p, 2)$
$f_U(3; p, 2)$	0.0967	0.1499	0.2018	0.2381	$f_U(2; 1-p, 2)$
$f_U(4; p, 2)$	0.0347	0.0691	0.1191	0.1786	$f_U(1; 1-p, 2)$
$f_U(5; p, 2)$	0.0062	0.0165	0.0391	0.0833	$f_U(0; 1-p, 2)$

$$f_U(a_r;\ {}_0,\ s) = f_U(2s - a_r + 1;\ -\lambda_0,\ s)$$

and consequently that

$$(3.10.5) \qquad f_U(a_r;\ p,\ s) = f_U(2s - a_r + 1;\ 1 - p,\ s).$$

This symmetry relation has been used together with the formulas just developed to compute the values of $f_U(a_r;\ p,\ s)$ shown in Table 2. The result for $p = 1/2$ in Table 2 is the general result for $p_\epsilon(t) = p_c(t)$.

3.11 Robustness of the Procedure with Respect to the Form of the Survival Curves. Box and Andersen [34], p. 1, state two properties which statistical criteria should possess; namely, they should (1) "be sensitive to change in the specific factors tested," and should (2) "be insensitive to changes, of a magnitude likely to occur in practice, in extraneous factors." The first property is related to the power of a test, and the second property they have called "robustness."

In our situation we are interested in the relative size of the medians of the control and experimental curves. We are not, however, particularly concerned with whether or not the two curves are of the logistic form. Thus the possible difference in medians is the "specific factor under test," and the forms of the curves are the "extraneous factors." We defer the discussion of the first property until a later paragraph, i.e., 3.13, since the power curve (or OC curve) of a test which is not robust is meaningless when the assumptions regarding the underlying distribution do not hold.

Our discussion of robustness goes beyond that of [34], in that we shall investigate the effect of a change in form of the survival curves on both α and β. Box and Andersen consider only the effect on α. We shall assume an underlying uniform distribution and compute approximations to actual probabilities of an error of the first kind, $\alpha^*(\alpha, \beta)$, and of the second kind, $\beta^*(\alpha, \beta)$, when the logistic analysis derived in paragraph 3.8 is applied. The $\alpha^*(\alpha, \beta)$ and $\beta^*(\alpha, \beta)$ will depend on the choice of p_0, p_1 and s, as well as the prescribed α and β. Since the linear survival curve, in a broad sense, departs further from the logistic than any sigmoidal curve we would expect to encounter in practice, small deviations of $\alpha^*(\alpha, \beta)$ and $\beta^*(\alpha, \beta)$ from α and β respectively would imply that the procedure based on the logistic model is robust.

We can find approximations to $\alpha^*(\alpha, \beta)$ and $\beta^*(\alpha, \beta)$ by using Wald's general method of deriving an approximation to the OC curve ([42], p. 158 and 160). We outline this method briefly. Let z, the test statistic, obey the following three assumptions:

(i) E(z) exists and is not equal to zero.

(ii) There exists a $\delta > 0$, such that $Pr(e^z < 1 - \delta) > 0$ and $Pr(e^z > 1 + \delta) > 0$.

(iii) For any real h, $E(e^{hz}) = g(h)$ exists.

Then, by Lemma A. 1. ([42], p. 158), there exists one and only one $h \neq 0$, such that

$$E(e^{hz}) = 1. \tag{3.11.1}$$

In general, the distribution of z depends on a parameter, say Θ, so that the solution to (3.11.1) is a function of Θ, i.e., $h = h(\Theta)$. Using this solution, we may write,[5]

$$(3.11.2) \qquad L(\Theta) \doteq \frac{A^{h(\Theta)} - 1}{A^{h(\Theta)} - B^{h(\Theta)}}$$

where $L(\Theta)$, the operating characteristic function, is the probability of acceptance when Θ is the true value of the parameter. If Θ_0 and Θ_1 are the values of Θ which define the zones of preference for acceptance and rejection, it follows that,

$$(3.11.3) \qquad \beta = L(\Theta_1) \doteq \frac{A^{h_1} - 1}{A^{h_1} - B^{h_1}}, \text{ where } h_1 = h(\Theta_1),$$

and

$$(3.11.4) \qquad \alpha = 1 - L(\Theta_1) \doteq \frac{1 - B^{h_0}}{A^{h_0} - B^{h_0}}, \text{ where } h_0 = h(\Theta_0).$$

In the usual case, $h_1 = -1$ and $h_0 = 1$, and $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$; so that (3.11.3) and (3.11.4) reduce to identities.

In our case, we have that $z(a_r) = \log \left[f_L(a_r; p_1, s)/f_L(a_r; p_0, s) \right]$,

---

[5]The symbol $\doteq$ denotes "is approximately equal to."

where a_r is distributed with frequency function, $f_U(a_r;\ p,\ s)$. Then (3.11.1) becomes

$$(3.11.5)\qquad \sum_{a_r=0}^{2s+1} \left[\frac{f_L(a_r;\ p_1,\ s)}{f_L(a_r;\ p_0,\ s)}\right]^{h(p_\gamma)} f_U(a_r;\ p_\gamma,\ s) = 1,\ \gamma = 0,\ 1.$$

For the values of p_0 and p_1 with which we deal, the three conditions cited above hold, and we can find the two pertinent solutions to (3.11.5); namely, $h_1 = h(p_1)$ and $h_0 = h(p_0)$. From (3.11.3) and (3.11.4), we have

$$(3.11.6)\qquad \beta^*(\alpha,\ \beta) \cdot\!= \frac{1 - \left(\frac{\beta}{1-\alpha}\right)^{h_1}}{\left(\frac{1-\beta}{\alpha}\right)^{h_1} - \left(\frac{\beta}{1-\alpha}\right)^{h_1}},$$

and

$$(3.11.7)\qquad \alpha^*(\alpha,\ \beta) \cdot\!= \frac{\left(\frac{1-\beta}{\alpha}\right)^{h_0} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{h_0} - \left(\frac{\beta}{1-\alpha}\right)^{h_0}}.$$

We have found by numerical methods the roots to (3.11.5) for $p_1 = 0.4$, $p_0 = 0.6$ and $p_1 = 0.3$, $p_0 = 0.7$; for $s = 0,\ 1,\ 2$. These roots are listed in Tables 3 and 4 along with the approximate values of $\alpha^*(\alpha,\ \beta)$ and $\beta^*(\alpha,\ \beta)$ for particular values of α and β found by using (3.11.6) and (3.11.7). Additional values of $\alpha^*(\alpha,\ \beta)$ and $\beta^*(\alpha,\ \beta)$ may be found by

Table 3

TABLE OF $\alpha^*(\alpha, \beta)$ AND $\beta^*(\alpha, \beta)$

FOR $p_0 = 0.6$, $p_1 = 0.4$; AND $s = 0, 1, 2$.

$s = 0$, $h_0 = -h_1 = 1.44$			
α	$\alpha^*(\alpha, \beta)$	β	$\beta^*(\alpha, \beta)$
0.05	0.014	0.05	0.014
0.10	0.041	0.10	0.041
0.20	0.120	0.20	0.120
0.05	0.016	0.10	0.039
0.10	0.039	0.05	0.016
0.10	0.045	0.20	0.109
0.20	0.109	0.10	0.045
$s = 1$, $h_0 = -h_1 = 1.22$			
0.05	0.027	0.05	0.027
0.10	0.064	0.10	0.064
0.20	0.156	0.20	0.156
0.05	0.028	0.10	0.063
0.10	0.063	0.05	0.028
0.10	0.067	0.20	0.149
0.20	0.149	0.10	0.067
$s = 2$, $h_0 = -h_1 = 1.15$			
0.05	0.038	0.05	0.038
0.10	0.074	0.10	0.074
0.20	0.169	0.20	0.169
0.05	0.034	0.10	0.073
0.10	0.073	0.05	0.034
0.10	0.077	0.20	0.164
0.20	0.164	0.10	0.077

Table 4

TABLE OF $\alpha^*(\alpha, \beta)$ AND $\beta^*(\alpha, \beta)$

FOR $p_0 = 0.7$, $p_1 = 0.3$; AND $s = 0, 1, 2$.

$s = 0$, $h_0 = -h_1 = 1.41$			
α	$\alpha^*(\alpha, \beta)$	β	$\beta^*(\alpha, \beta)$
0.05	0.016	0.05	0.016
0.10	0.043	0.10	0.043
0.20	0.124	0.20	0.124
0.05	0.016	0.10	0.040
0.10	0.040	0.05	0.016
0.10	0.047	0.20	0.114
0.20	0.114	0.10	0.047
$s = 1$, $h_0 = -h_1 = 1.17$			
0.05	0.031	0.05	0.031
0.10	0.071	0.10	0.071
0.20	0.165	0.20	0.165
0.05	0.032	0.10	0.070
0.10	0.070	0.05	0.032
0.10	0.074	0.20	0.159
0.20	0.159	0.10	0.074
$s = 2$, $h_0 = -h_1 = 1.115$			
0.05	0.036	0.05	0.036
0.10	0.079	0.10	0.079
0.20	0.176	0.20	0.176
0.05	0.037	0.10	0.078
0.10	0.078	0.05	0.037
0.10	0.082	0.20	0.172
0.20	0.172	0.10	0.082

substituting the tabled values of h_0 and h_1 in (3.11.6) and (3.11.7). Two symmetry properties which hold when $p_0 = 1 - p_1$ were found useful in computing these tables. These are:

(1) $h_0 = - h_1$, and

(2) $\alpha^*(\alpha, \beta) = \beta^*(\beta, \alpha)$.

The first relation follows directly from applying to (3.11.5) the symmetry properties of $f_L(a_r; p, s)$, i.e., (3.8.8), and of $f_U(a_r; p, s)$, i.e., (3.10.5). The second follows from using the first relation in (3.11.6) and (3.11.7).

The results of Tables 3 and 4 point up the relative insensitivity of the procedure to the form of the survival curve. In fact, they indicate that the logistic model is conservative relative to the uniform with respect to the probabilities of the two types of errors. Further as s increases, the deviations between the prescribed and actual probabilities decrease; and when p_0 and p_1 are chosen farther apart these deviations also decrease slightly. Though the results of Tables 3 and 4 pertain to cases in which the p_0 and p_1 are chosen symmetrically about one-half, additional calculations indicate that an asymmetric choice yields results close to those tabled.

The curves which we considered in paragraph 3.6 were sigmoidal with a range which had a finite nonnegative lower bound and an infinite upper bound. This kind of range is more realistic than either the uniform, which has a

finite range, or the logistic, which has a range infinite in both directions. Intuitively, it seems that for these underlying distributions the true α will fall between the α and $\alpha^*(\alpha, \beta)$ shown in Tables 3 and 4 and the true β will fall between β and $\beta^*(\alpha, \beta)$. Thus it seems that the analysis based on the logistic model is slightly conservative when the underlying distributions are of the form of those discussed in paragraph 3.6.

Although this procedure is robust when the underlying distributions are uniform, results for a particular case indicate that the underlying sample exponential distributions will yield probabilities of errors greatly different from their nominal values. This is not surprising, since the exponential curve is uniformly concave. Thus, particularly in the vicinity of the median, the exponential is worse than the uniform as an approximation to the logistic.

3.12 Comparison of the Original Analysis to the Logistic. The test statistic proposed in [38] is

$$z(a_r) = \log \left(\frac{p_1}{p_0}\right)^x \left(\frac{1-p_1}{1-p_0}\right)^y , \tag{3.12.1}$$

where $x = \frac{a_r(s+1)}{2s+1}$, $y = \frac{s(2s-a_r+1)}{2s+1}$, and $a_r = 0, 1, \ldots, 2s+1$. In the particular case treated in [38], $p_0 = 1 - p_1 = 0.6$ and $s = 2$, so that (3.12.1) yields

$$z(0) = 0.3522$$

$$z(1) = 0.1761$$

$$z(2) = 0.0000$$

$$z(3) = -0.1761$$

$$z(4) = -0.3522$$

$$z(5) = -0.5283.$$

The prescribed probabilities of errors were $\alpha = \beta = 0.10$.

Using the methods of the previous paragraph, we can find the approximations to the true probabilities of errors, i.e., $\alpha'(0.10, 0.10)$ and $\beta'(0.10, 0.10)$, when the underlying distributions are of the logistic form. Obviously the three conditions on the $z(a_r)$ hold when a_r is distributed with frequency function, $f_L(a_r; p, s)$, so there exist unique solutions to the equations,

$$\sum_{a_r=0}^{5} e^{z(a_r)h(p_\gamma)} f_L(a_r; p_\gamma, 2) = 1,$$

where $\gamma = 0, 1$, and $p_0 = 1 - p_1 = 0.6$. These solutions are $h(0.6) = 2.43$ and $h(0.4) = 0.16$. Substituting these results in (3.11.3) and (3.11.4), we have

$$\alpha'(0.10, 0.10) = \frac{1}{1 + 9^{2.43}} = 0.004,$$

and

$$\beta'(0.10, 0.10) = \frac{9^{0.16}}{1 + 9^{0.16}} = 0.587.$$

Thus, when the logistic model holds and the original analysis is used, the probability of rejecting an experimental compound with $p_\epsilon(m_c) = 0.6$ is 0.004 and the probability of accepting an experimental compound with $p_\epsilon(m_c) = 0.4$ is 0.587. One might think that a great number of errors of the second kind would result from using this analysis. However, this is not necessarily the case. In the application cited in [38], most of the compounds tested had very poor protective properties, so that very few were accepted. Thus, relatively few errors of the second kind were made and the apparent gross underestimate of the true β did not result in an excessive amount of testing on poor compounds. In other types of screening experiments, however, the underestimation of β might be more serious.

Sometimes, in this experiment, a compound which had been rejected at the screening stage was later subjected to more exhaustive tests; usually because it belonged to a family containing one or more known protective compounds. In no case did such a compound prove to have useful protective properties. This is reasonable in view of the apparent small magnitude of the true value of α.

3.13 The OC Function. The operating characteristic (OC) function, $L(p)$, is the probability of acceptance when $p_\epsilon(m_c) = p$. Wald [42], p. 160, cites the following approximation to $L(p)$,

$$L(p) \doteq \frac{A^{h(p)} - 1}{A^{h(p)} - B^{h(p)}}, \tag{3.13.1}$$

where $h(p) \neq 0$ is the unique solution to the equation,

$$(3.13.2) \qquad \sum_{a_r=0}^{2s+1} \left[\frac{f_L(a_r;\ p_1,\ s)}{f_L(a_r;\ p_0,\ s)} \right]^{h(p)} f_L(a_r;\ p,\ s) = 1.$$

This approximation is subject to the three conditions cited in paragraph 3.11.

When $p = p'$ such that $E_{p'}(z) = 0$, that is, when condition (i) cited in paragraph 3.11 does not hold; we have (see [42], p. 176), that,

$$(3.13.3) \qquad L(p') \doteq \frac{\log A}{\log A - \log B} .$$

As before, to evaluate (3.13.3), we let $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$.

We have considered the case where $p_0 = 1 - p_1 = 0.6$ and $\alpha = \beta = 0.10$, $s = 0, 1, 2$. This situation permits several simplifications in using the above formulas. Because of the symmetry property, (3.8.8), $h(p) = - h(1 - p)$ and consequently $L(p) \doteq 1 - L(1 - p)$. Here $p' = 1/2$, so that $L(1/2) = 1/2$. These results hold whenever $p_0 = 1 - p_1$ and $\alpha = \beta$. The approximations to $L(p)$ are shown for intervals of one-tenth in Table 5. No differences among the $h(p)$ for $s = 0, 1, 2$, were detected at the level of accuracy attained in these calculations. The values of $h(p)$ shown here may be used for computing $L(p)$ for other values of α and β. A graph of $L(p)$ was made from Table 5, and appears in Figure 1.

These results show that the procedure is quite powerful. Further, it

Table 5

TABLE OF THE OC AND ASN FUNCTION

FOR $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; AND $s = 0, 1, 2$

			$E_p(n)$		
p	h(p)	L(p)	s = 0	s = 1	s = 2
0.0	$-\infty$	0.000	8.1	3.2	2.0
0.1	- 5.5	0.000	12.9	4.4	2.7
0.2	- 3.44	0.001	18.7	6.2	3.7
0.3	- 2.10	0.010	28.9	9.5	5.7
0.4	- 1.00	0.100	48.3	15.7	10.4
0.5	0.00	0.500	66.0	21.4	12.7
0.6	1.00	0.900	48.3	15.7	10.4
0.7	2.10	0.990	28.9	9.5	5.7
0.8	3.44	0.999	18.7	6.2	3.7
0.9	5.5	1.000	12.9	4.4	2.7
1.0	∞	1.000	8.1	3.2	2.0

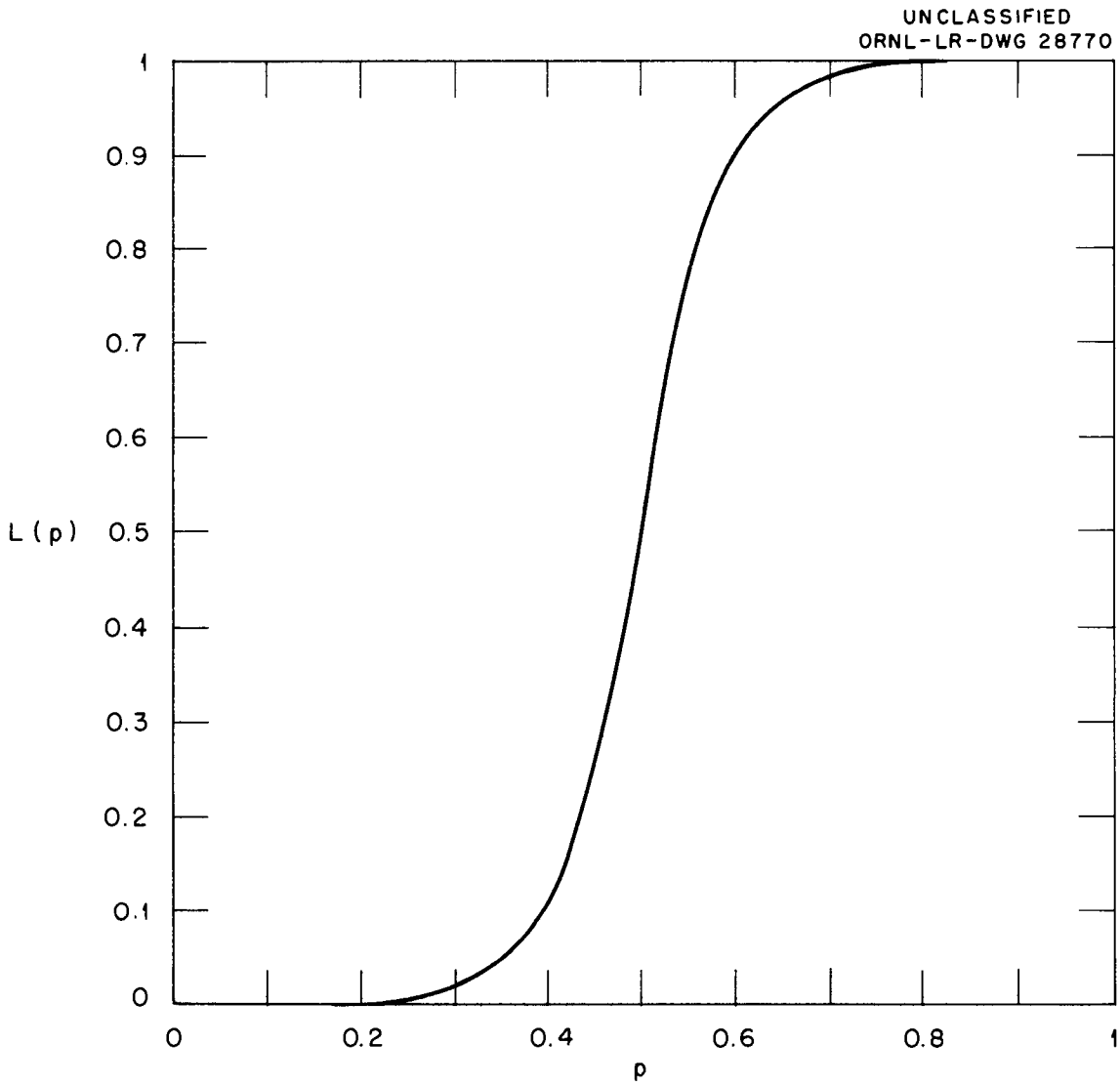

Figure 1.
The OC Function for
$p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$
and $s = 0, 1, 2$.

apparently possesses the requirements concerning the OC function cited in [42], p. 32; namely that,

$$1 - L(p) \leq \alpha \text{ for all } p \geq p_0,$$
$$L(p) \leq \beta \text{ for all } p \leq p_1 .$$

We have shown that those requirements hold for the particular case considered to the order of approximation of (3.13.1). Intuitively, it seems that this should be true in general. If these requirements did not hold for a particular choice of p_0 and p_1, an optimum weight function would have to be found, and the entire procedure would be considerably more complicated.

3.14 The ASN Function. The average sample number (ASN) function, $E_p(n)$, is the average number of trials needed to reach a decision. When the three conditions cited in paragraph 3.11 hold, Wald gives the approximate formula (see [42], p. 172),

$$(3.14.1) \qquad E_p(n) \doteq \frac{L(p) \log B + [1 - L(p)] \log A}{E_p [z(a_r)]} ,$$

in which the usual approximations for A and B may be used.

When $p = p'$, such that $E_{p'} [z(a_r)] = 0$, an approximation to $E_{p'}(n)$ is given by (see [42], p. 176),

$$(3.14.2) \qquad E_{p'}(n) = - \frac{\log B \log A}{E_{p'}\left[z^2(a_r)\right]} .$$

We have used (3.14.1) and (3.14.2) together with the results for $L(p)$ shown in Table 5 to approximate $E_p(n)$ for the cases where $p_0 = 1 - p_1 = 0.6$, $\alpha = \beta = 0.10$, and $s = 0, 1, 2$. The numerical results appear in Table 5 and a graph made from those values appears as Figure 2. The symmetry relation,

$$(3.14.3) \qquad E_p\left[z(a_r)\right] = - E_{1-p}\left[z(a_r)\right]$$

where $p_0 = 1 - p_1$, was found useful in computing these results.

An examination of Table 5 shows the following interesting approximate relation:

$$(3.14.4) \qquad E_p(n)\Big|_{s=0} \doteq 3E_p(n)\Big|_{s=1} \doteq 5E_p(n)\Big|_{s=2} .$$

In other words, the average total number of mice used appears to be independent of the group size for $s = 0, 1, 2$. This is in contrast to the double dichotomy, where using larger groups always increases the $E_p(n)$ (see [42], p. 102 and p. 116).

The question arises as to what group size is optimum with respect to some statistical criterion. In formulating a criterion we may neglect the OC function, since it is apparently insensitive to group size, for

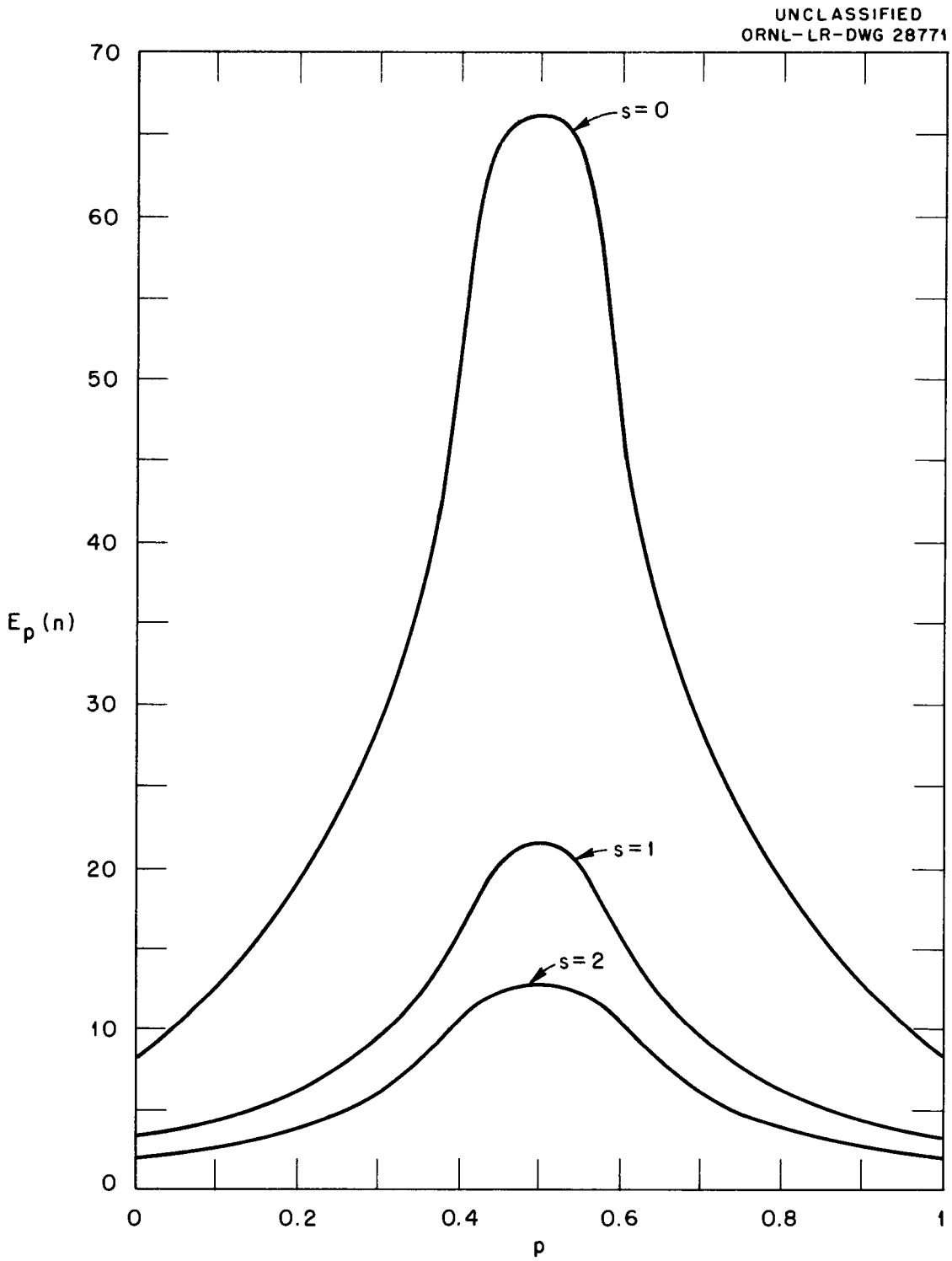

Figure 2.
The ASN Function for
$p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$
and $s = 0, 1, 2$.

$s = 0, 1, 2$; and also the average total number of mice used, since this is apparently invariant for $s = 0, 1, 2$. Thus, we may define an optimum group size as one which minimizes the average time needed to reach a decision, that is, s is said to be optimum if it minimizes $E_p(n) \cdot E_p(t)$. Now for the logistic survival curve $E_p(t)$ is equal to μ_c for all s, so that this criterion reduces to finding the s for which $E_p(n)$ is a minimum. For $s = 0, 1, 2$, group sizes of five ($s = 2$) achieve a minimum uniformly in p. However, it is not clear whether the largest feasible group size is always optimum according to this criterion. Further investigation of the criterion for various values of p_0 and p_1 as well as for larger group sizes would be needed. Even if it were proved that the largest feasible group size were always optimum, considerations other than statistical might dictate a smaller group size.

As we have noted, using the usual approximations for A and B decreases the probabilities of the errors, and consequently increases the $E_p(n)$. The method of Wald [42], p. 65, could be used to find closer approximations to $E_p(n)$ than those given in Table 5.

3.15 The Effect of Truncation. Although Wald [42], pp. 151-152, has shown that the probability is one that the test procedure will eventually terminate, it is sometimes desirable to set a definite upper limit, say n_0, on the number of trials. If, by the usual rules, no decision is reached by the n_0^{th} trial, we reject if $\sum_{r=1}^{n_0} z(a_r) > 0$ and we accept if $\sum_{r=1}^{n_0} z(a_r) \leq 0$. This truncation increases the probabilities of errors of the first and second kinds from the nominal values to values which we denote by $\alpha(n_0)$ and $\beta(n_0)$.

Wald [42], pp. 63-64, has derived a general method of deriving upper bounds for $\alpha(n_0)$ and $\beta(n_0)$; namely,

$$\alpha(n_0) \leq \alpha + \Phi(\gamma_2) - \Phi(\gamma_1), \tag{3.15.1}$$

where

$$\gamma_1 = - \frac{\sqrt{n_0}\, E_0\left[z(a_r) \right]}{\sigma_0\left[z(a_r) \right]}, \tag{3.15.2}$$

and

$$\gamma_2 = \frac{\log A - n_0\, E_0\left[z(a_r) \right]}{n_0\, \sigma_0\left[z(a_r) \right]} ; \tag{3.15.3}$$

$$\beta(n_0) \leq \beta + \Phi(\gamma_4) - \Phi(\gamma_3) , \tag{3.15.4}$$

where

$$\gamma_3 = \frac{\log B - n_0\, E_1\left[z(a_r) \right]}{\sqrt{n_0}\, \sigma_1 \left[z(a_r) \right]} , \tag{3.15.5}$$

and

$$\gamma_4 = - \frac{n_0\, E_1 \left[z(a_r) \right]}{\sqrt{\sigma_1} \left[z(a_r) \right]} . \tag{3.15.6}$$

E_i and σ_i denote the expectation and standard deviation, respectively, when $p = p_i$, $i = 0, 1$.

Table 6

UPPER BOUNDS ON $\alpha(n_0)$ AND $\beta(n_0)$

FOR $p_0 = 0.6$, $p_1 = 0.4$; $\alpha = \beta = 0.10$; AND $s = 0, 1, 2$.

s = 0		s = 1		s = 2	
n_0	Upper Bound on $\alpha(n_0)$, $\beta(n_0)$	n_0	Upper Bound on $\alpha(n_0)$, $\beta(n_0)$	n_0	Upper Bound on $\alpha(n_0)$, $\beta(n_0)$
3	0.591	1	0.591	-	-
5	0.481	-	-	1	0.477
10	0.433	-	-	2	0.427
15	0.396	5	0.392	3	0.390
25	0.339	-	-	5	0.332
30	0.316	10	0.312	6	0.310
45	0.265	15	0.261	9	0.259
50	0.252	-	-	10	0.245
60	0.229	20	0.228	12	0.223
75	0.203	-	-	15	0.197
100	0.173	-	-	20	0.167
198	0.122	64	0.122	38	0.121

The results of truncation of the procedure when $p_0 = 1 - p_1 = 0.6$ and $s = 0, 1, 2$, for various values of n_0 are shown in Table 6. Because of the symmetry relation (3.14.3), the relation,

$$\sigma_0 \left[z(a_r) \right] = \sigma_1 \left[z(a_r) \right]$$

and the fact that $\log A = - \log B$, we have $\gamma_4 = - \gamma_1$ and $\gamma_3 = - \gamma_2$. Thus by the summetry of the normal distribution the bounds given by (3.15.1) and (3.15.4) are identical.

Table 6 has been arranged such that the total number of mice used is n_0 trials is equal for $s = 0, 1, 2$. An inspection of Table 6 shows that the effects of truncation apparently depend on the total number of mice used and not on the group size, at least for $s = 0$, 1, and 2. The last row of entries in Table 6 is for $n_0 = 3 \operatorname{Max}_p E_p(n)$. Wald recommends this choice of n_0, stating that, in general, it will not affect α and β greatly. The statement is certainly verified in Table 6. However, for screening tests this choice of n_0 is usually impracticable.

3.16 An Illustration of the Procedure. We shall apply the logistic analysis developed in paragraph 3.8 to the data collected by Doherty, Burnett, and Shapira [36]. Each trial was performed as follows: Ten male mice of the C_3H strain, each ten to twelve-weeks old and each weighing twenty to twenty-four grams, were randomly divided into two groups of five. One group was injected intraperitoneally with a saline solution and one with the experimental compound. In this instance, each mouse of the experimental group was treated with 30.0 mg (190.0 μM) of cysteine hydrochloride.

Ten minutes after injection the ten mice were simultaneously exposed to 800 r of X-radiation. The two groups were observed until the third mouse in the control died. The number of mice surviving in the experimental group was then counted, and the decision was made to accept, reject or continue sampling. The trials were continued sequentially until a decision was reached.

In terms of our notation, $s = 2$ and the probabilities of errors were chosen to be $\alpha = \beta = 0.10$ with the zones of preference determined by $p_0 = 1 - p_1 = 0.6$. Using Table 1, we find

$$z(0) = -z(5) = \log f_L(0;\ 0.4,\ 2) - \log f_L(0;\ 0.6,\ 2) = 0.4806$$
$$z(1) = -z(4) = \log f_L(1;\ 0.4,\ 2) - \log f_L(1;\ 0.6,\ 2) = 0.2882$$
$$z(2) = -z(3) = \log f_L(2;\ 0.4,\ 2) - \log f_L(2;\ 0.6,\ 2) = 0.0961.$$

Further $\ln(\alpha/1 - \beta) = -\ln(\beta/1 - \alpha) = 0.9542$. The results of the decision procedure for cysteine hydrochloride are summarized in Table 7.

Table 7

The Decision Procedure for Cysteine Hydrochloride

r	a_r	$z(a_r)$	$\Sigma z(a_r)$	Decision
1	3	-0.0961	-0.0961	Since $-0.9542 < -0.0961 < 0.9542$ continue sampling
2	5	-0.4806	-0.5767	Since $-0.9542 < -0.5761 < 0.9542$ continue sampling
3	5	-0.4806	-1.0573	Since $-1.0573 < -0.9542$, accept the compound

3.17 Possible Extensions. The most obvious need for further work on this problem is with regard to the computing of tables of $f_L(a_r; p, s)$ for larger values of s. Probably the most practicable way of accomplishing this is to use (3.8.5) in conjunction with a high speed computer. Such tables, together with similar ones for $f_U(a_r; p, s)$, would permit further studies to be made on the robustness of the procedure. Then the ASN function could be computed for further values of s, and the generality of the approximate relation (3.14.4) (concerning the invariance of the average total number of mice used) could be ascertained.

The possibility of a stopping time based on an order statistic other than the median could be considered. As was pointed out in paragraph 3.3, it is thought that this will lead to a procedure less powerful than the one we have considered. The ASN function might also be affected adversely.

Probably the easiest and most useful extension is to apply this sampling method to nonsigmoidal survival curves, in particular, to the simple exponential. Preliminary work on this indicates that this generalization is quite simple and leads to numerical results differing appreciably from those for underlying logistic distributions.

3.18 Summary. Part III investigates methods of analyzing data collected by using the sampling procedure proposed by Kimball, Burnett, and Doherty [38]. The method of analysis which is based on logistic survival curves was found to be best with regard to practical and theoretical considerations. It is shown that the procedure is relatively insensitive to the form of the sigmoidal curves assumed. The OC function, the ASN function, and the effects of truncation are investigated for particular cases. An illustration of the procedure is given.

IV. ACKNOWLEDGMENTS

I wish to express my appreciation:

To the National Science Foundation and the Oak Ridge Institute of Nuclear Studies for providing the fellowships which enabled me to write this thesis;

To Mrs. Charlotte Weaver, Mrs. Edna Jones and Mrs. Nette G. Crowe for their careful preparation of the final manuscript;

To Professors J. E. Freund and C. W. Clunies-Ross for their criticisms and assistance in the writing of Part I;

To Dr. D. A. Gardiner for his helpful criticisms of Part III;

To Professor R. A. Bradley for his careful formulation of the problem in Part II, as well as his helpful suggestions with regard to both Part II and Part III;

To Dr. A. W. Kimball for his careful formulation of the problem in Part III, his invaluable guidance during the progress of the work, and his critical reading of the final draft;

To Professor Boyd Harshbarger for his continuing encouragement and generosity toward a sometimes cantankerous graduate student.

V. BIBLIOGRAPHY

Literature Cited in I.

1. Aitken, A. C. and Silverstone, N., "On the estimation of statistical parameters," Proceedings of the Royal Society of Edinburgh, Section A, 61 (1942), 186-194.

2. Barankin, E. W., "Concerning some inequalities in the theory of statistical estimation," Skandinavisk Aktuarietidskrift, 34 (1951), 35-40.

3. Barankin, E. W., "Locally best unbiased estimates," The Annals of Mathematical Statistics, 20 (1949), 477-501.

4. Bhattacharyya, A., "On some analogues to the amount of information and their uses in statistical estimation," Sankhya, 8 (1946), 1-14.

5. Chapman, D. G. and Robbins, H., "Minimum variance estimation without regularity conditions," The Annals of Mathematical Statistics, 22 (1951), 581-586.

6. Cramér, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

7. Cramér, H., "Contributions to the theory of statistical estimation," Skandinavisk Aktuarietidskrift, 29 (1946), 85-94.

8. Eisenhart, C., "The assumptions underlying the analysis of variance," Biometrics, 3 (1947), 1-26.

9. Hammersley, J., "On estimating restricted parameters," Journal of the Royal Statistical Society, Series B, 12 (1950), 192-240.

10. Koopman, B. O., "On distributions admitting a sufficient statistic," Transactions of the American Mathematical Society, 39 (1936), 399-409.

11. Pitman, E. J. G., "Sufficient statistics and intrinsic accuracy," Proceedings of the Cambridge Philosophical Society, 32 (1936), 567-579.

12. Rao, C. R., "Information and accuracy attainable in the estimation of statistical parameters," Bulletin of the Calcutta Mathematics Society, 37 (1945), 81-91.

13. Tweedie, M. C. K., "Functions of a statistical variate with given reference to Laplacian distributions," Proceedings of the Cambridge Philosophical Society, 43 (1947), 41-49.

14. Tweedie, M. C. K., Notes on Generating Functions and Distribution Theory, Virginia Polytechnic Institute, 1953.

15. Wald, A., Statistical Decision Functions, New York, John Wiley and Sons, Inc., 1950.

16. Widder, D. V., The Laplace Transform, Princeton, Princeton University Press, 1946.

17. Wilks, S. S., Mathematical Statistics, Princeton, Princeton University Press, 1950.

18. Wolfowitz, J., "The efficiency of sequential estimates and Wald's equation for sequential processes," The Annals of Mathematical Statistics, 18 (1947), 215-230.

Literature Cited in II

19. Bradley, R. A., "Rank analysis of incomplete block designs. III. Some large results on estimation and power for a method of paired comparisons," Biometrika, 42 (1955), 450-470.

20. Bradley, R. A., and Terry, M. E., "Rank analysis of incomplete block designs. I. The method of paired comparisons," Biometrika 39 (1952), 324-345.

21. Chanda, K. C., "A note on the consistency and maxima of the roots of likelihood equations," Biometrika, 41 (1954), 56-61.

22. Chiang, C. L., "On regular best asymptotically normal estimates," The Annals of Mathematical Statistics, 27 (1956), 336-351.

23. Cramér, H., Random Variables and Probability Distributions, Cambridge, Cambridge University Press, 1946.

24. Cramér, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1937.

25. Dugué, E., "Applications des propriétés de la limite au sens du calcul des probabilités a l'étude de diverses questions d'estimation," Journal de l'École Polytechnique, 3 (1937), 305-

26. Dykstra, O., "Rank analysis of incomplete block designs; Unequal numbers of repetitions of a method of paired comparisons," unpublished.

27. Fisher, R. A., "On an absolute criterion for fitting frequency curves," Messenger of Mathematics, 41 (1912), 155-160.

28. Huzurbazar, V. S., "The likelihood equation, consistency, and maxima of the likelihood function," The Annals of Eugenics, 14 (1948), 185-200.

29. Kraft, C. and LeCam, L., "A remark on the roots of the maximum likelihood equation," The Annals of Mathematical Statistics, 27 (1956), 1174-1177.

30. Neyman, J. and Scott, E. L., "Consistent estimates based on partially consistent observations," Econometrica, 16 (1948), 1-32.

31. Noether, G. E., "On a theorem of Pitman," The Annals of Mathematical Statistics, 26 (1955), 64-68.

32. Wald, A., "Tests of statistical hypotheses concerning several parameters when the number of observations is large," Transactions of the American Mathematical Society, 54 (1943), 426-482.

33. Wald, A., "Note on the consistency of the maximum likelihood estimate," The Annals of Mathematical Statistics, 20 (1949), 595-601.

Literature Cited in III.

34. Box, G. E. P. and Andersen, S. L., "Permutation theory in the derivation of robust criteria and the study of departures from assumption," Journal of the Royal Statistical Society, Series B, 17 (1955), 1-34.

35. Cramer, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

36. Doherty, D. G., Burnett, W. T., and Shapira, R., "Chemical protection against ionizing radiation II. Mercaptoalkyamines and related compounds with protective activity," Radiation Research, 7 (1957), 13-21.

37. Epstein, B. and Sobel, M., "Sequential tests in the exponential case," The Annals of Mathematical Statistics, 26 (1955), 82-93.

38. Kimball, A. W., Burnett, W. T., and Doherty, D. G., "Chemical protection against ionizing radiation I. Sampling methods for screening compounds in radiation protection studies with mice," Radiation Research, 7 (1957), 1-12.

39. MacRobert, T. M., Functions of a Complex Variable, London, MacMillan and Company, Ltd., 1950.

40. Pearson, K., Tables of the Incomplete Beta Function, Cambridge, Cambridge University Press, 1934.

41. Rao, C. R., "Sequential tests of null hypotheses," Sankhya, 10 (1950), 361-370.

42. Wald, A., Sequential Analysis, New York, John Wiley and Sons, Inc., 1947.

43. Wald, A., and Wolfowitz, J., "Optimum character of the probability ratio test," The Annals of Mathematical Statistics, 19 (1948), 326-339.

44. Winsor, C. P., "A comparison of certain symmetrical growth curves," Journal of the Washington Academy of Sciences, 22 (1932), 73-84.

Literature Examined for II.

Ogawa, J. and Moustafa, M. D. and Roy, S. N., On the Asymptotic Distribution of the Likelihood Ratio in Some Problems on Mixed Variate Populations, Chapel Hill, N. C., Institute for Statistics, 1957.

Wald, A., "Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process," The Annals of Mathematical Statistics, 19 (1948), 40-46.

Wilks, S. S., "The large sample distribution of the likelihood ratio for testing composite hypotheses," The Annals of Mathematical Statistics, 9 (1938), 60-62.

Literature Examined for III

Albert, G. E., "On the computation of the sampling characteristics of a general class of sequential decision problems," The Annals of Mathematical Statistics, 25 (1954), 340-356.

Erdeyli, A., Higher Transcendental Functions, Vol. 1, New York, McGraw-Hill Book Company, Inc., 1953, 56 ff.

Girshick, M. A., "An extension of the optimum property for the sequential probability ratio test," The Annals of Mathematical Statistics, 29 (1958), 288-290.

VI. APPENDICES

Appendix A. The Covariance in Terms of Conditional Expectations.

Let $U = (u_1, u_2, \ldots, u_i)$ and $V = (v_1, v_2, \ldots, v_j)$ be random variables. Assume $p = p(U, V)$ and $q = q(U, V)$ have finite means and variances. Then we have

$$\text{(i)} \quad \mathrm{Cov}(p, q) = EE(p.q|V) - EE(p|V)\, EE(q|V) \ .$$

But,

$$\mathrm{Cov}(p, q|V) = E(p.q|V) - E(p|V)\, E(q|V) \ ,$$

from which, taking expectations with respect to V, we have,

$$E\Big[\, \mathrm{Cov}(p, q|V) \,\Big] = EE(p\ q|V) - E\Big[\, E(p|V)\, E(q|V) \,\Big] \ .$$

Substituting this result in (i) gives

$$\text{(ii)} \quad \mathrm{Cov}(p, q) = E\Big[\mathrm{Cov}(p, q|V)\Big] + E\Big[E(p|V).E(q|V)\Big] - EE(p|V)\, EE(q|V) \ ,$$

$$\text{(iii)} \quad \mathrm{Cov}(p, q) = E\Big[\mathrm{Cov}(p, q|V)\Big] + \mathrm{Cov}\Big[E(p|V), E(q|V)\Big]. \quad \text{Q.E.D.}$$

Appendix B. A Complex Integration Theorem.

Theorem (cited without proof by MacRobert [39], p. 74, problem 37). Let $P(z)$ and $Q(z)$ be polynomials of degree m and n respectively, where $m \leq n - 2$, and let $Q(z)$ have no positive or zero real roots. Then,

$$\int_0^\infty \frac{P(x)}{Q(x)}\, dx = - \mathcal{R} \ ,$$

where $\mathcal{R}$ denotes the sum of the residues of $P(z)\ln z/Q(z)$, $(0 < \text{amp } z < 2\pi)$ at the zeros of $Q(z)$.

Proof: Integrate $f(z) = P(z) \ln z/Q(z)$ around the contour of Figure 3, consisting of:

(1) the x-axis from r to R ,

(2) the large circle $|z| = R$,

(3) the x-axis from R to r,

(4) the small circle $|z| = r$.

Consider the integral along (4) which as $r \longrightarrow 0$, is $2\pi i$ times the residue of $f(z)$ at $z = 0$. By a theorem cited in [39], p. 57, this integral is equal to $- 2\pi i \lim_{z \to 0} z\, f(z)$, if this limit exists. We may write,

$$\text{(i)} \quad \lim_{z \longrightarrow 0} |z\, f(z)| = \lim_{z \longrightarrow 0} \frac{|z|\ |\ln z|\ \left|\sum_{k=0}^{m} A_k z^k\right|}{\left|\sum_{j=0}^{n} B_j z^j\right|} ,$$

where $P(z) = \sum_{k=0}^{m} A_k z^k$ and $Q(z) = \sum_{j=0}^{n} B_j z^j$. Now, on (4), we have,

(ii) $|z| = r$,

(iii) $|\ln z| = |\ln r + i\theta| \leq |\ln r| + \theta$, where $0 \leq \theta < 2\pi$,

(iv) $\left|\sum_{k=0}^{m} A_k z^k\right| \leq \sum_{k=0}^{m} |A_k| r^k$, and

(v) $\left|\sum_{j=0}^{n} B_j z^k\right| \geq |B_0| - \sum_{j=1}^{n} |B_j| r^j$.

When r is small enough the expression on the right hand side of (v) is positive, since $B_0 \neq 0$. Applying (ii) - (v) to (i), we have

UNCLASSIFIED
ORNL-LR-DWG 28772

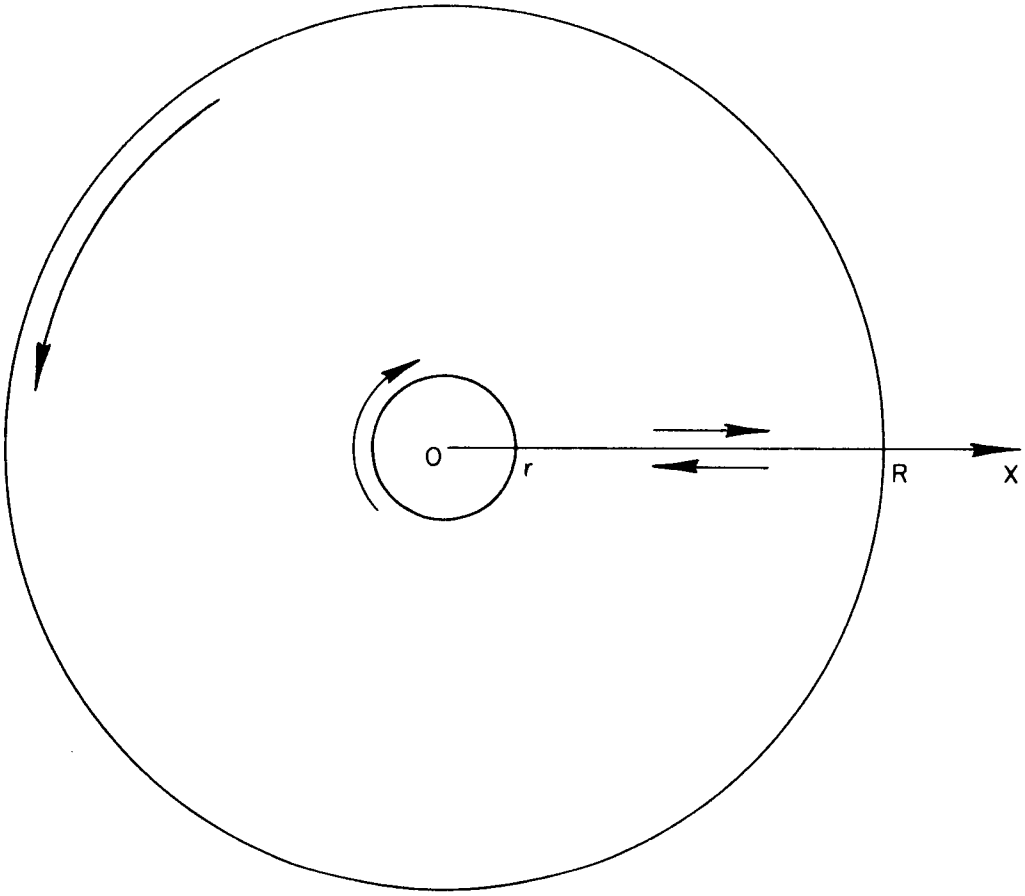

Figure 3.
The Contour of Integration.

$$\lim_{z \to 0} |z\, f(z)| \leq \lim_{r \to 0} \left\{ \frac{r(|\ln r| + \theta) \sum_{i=0}^{m} |A_i| r^i}{|B_0| - \sum_{j=1}^{n} |B_j| r^j} \right\} .$$

In the limit, the summation in the denominator goes to zero and the summation in the numerator to $|A_0|$; so that if we assume $A_0 \neq 0$, we may write,

$$\lim_{z \to 0} |z\, f(z)| \leq \frac{|A_0|}{|B_0|} \lim_{r \to 0} \left\{ \frac{(|\ln r| + \theta)}{(1/r)} \right\} .$$

Applying l'Hôpital's rule, we have

$$\lim_{z \to 0} |z\, f(z)| \leq \frac{|A_0|}{|B_0|} \lim_{r \to 0} (-r) = 0.$$

When $A_0 = 0$, the inequality follows <u>a fortiori</u>. Therefore $\lim_{z \to 0} \left\{ -z\, f(z) \right\} = 0$, and the integral along (4) of $f(z)$ is zero as $r \to 0$.

Consider now the integral along (2), which, as $R \to \infty$, is the negative of $2\pi i$ times the residue of $f(z)$ at infinity. By [39], p. 58, this integral is equal to $2\pi i$. $\lim_{z \to \infty} \left\{ z\, f(z) \right\}$, if this limit exists. Proceeding as above, we may write,

$$\text{(vi)} \quad \lim_{z \to \infty} |z\, f(z)| = \lim_{z \to \infty} \left\{ \frac{|z|\ |\ln z|\ \left| \sum_{i=0}^{m} A_k z^k \right|}{\left| \sum_{j=0}^{n} B_j z^j \right|} \right\} .$$

On (2), we have

(vii) $|z| = R$,

(viii) $|\ln z| = |\ln R + i\theta| \leq |\ln R| + \theta$, where $0 \leq \theta < 2\pi$,

(ix) $\left| \sum_{i=0}^{m} A_k z^k \right| \leq \sum_{i=0}^{m} |A_k| R^k$, and

(x) $\left| \sum_{j=0}^{n} B_j z^j \right| \geq |B_n| R^m - \sum_{j=0}^{n-1} |B_j| R^j$.

When R is large enough the expression on the right of (x) is positive, since $B_n \neq 0$. Applying (vii) - (x) to (vi), we may write,

$$\lim_{z \to \infty} |z\, f(z)| \leq \lim_{R \to \infty} \left\{ \frac{(|\ln R| + \theta) \sum_{k=0}^{m} |A_k| R^{k+1}}{|B_n| R^n - \sum_{j=0}^{n-1} |B_j| R^j} \right\} .$$

Dividing numerator and denominator by R^n, we have

$$\lim_{z \to \infty} |z\, f(z)| \leq \lim_{R \to \infty} \left\{ \frac{(\ln R + \theta) \sum_{k=0}^{m} |A_k| R^{k-n+1}}{|B_n| - \sum_{j=0}^{n-1} |B_j| R^{j-n}} \right\} .$$

In the limit the denominator goes to $|B_n|$, so we may write,

$$\lim_{z \to \infty} |z\, f(z)| \leq \frac{1}{|B_n|} \lim_{R \to \infty} \left\{ \frac{(|\ln R| + \theta)}{\left(\sum_{k=0}^{m} |A_k| R^{k-n+1} \right)^{-1}} \right\} .$$

Since $n > m + 1$, the denominator becomes infinite in the limit and we may apply l'Hôpital's rule, which yields,

$$\lim_{z \to \infty} |z\, f(z)| \leq \frac{1}{|B_n|} \lim_{R \to \infty} \left\{ \frac{\left(\sum_{k=0}^{m} |A_k|\, R^{k+1-n} \right)^2}{R\left(\sum_{k=0}^{m} |A_k|\, (n-k-1) R^{k-n} \right)} \right\} .$$

Multiplying numerator and denominator by R^{2n}, we have

$$\lim_{z \to \infty} |z\, f(z)| \leq \frac{1}{|B_n|} \lim_{R \to \infty} \left\{ \frac{\left(\sum_{k=0}^{m} |A_k|\, R^{k+1} \right)^2}{\left(\sum_{k=0}^{m} |A_k|\, (n-k+1)\, R^{n+k+1} \right)} \right\} .$$

Since the denominator is an infinity of higher order than the numerator, the expression on the right goes to zero, and the integral along (2) $\longrightarrow 0$ as $R \longrightarrow \infty$.

Therefore, as $r \longrightarrow 0$ and $R \longrightarrow \infty$, the integral along the entire contour may be written as

$$2\pi i\, \mathcal{R} = \int_{(1)} \frac{P(z) \ln z\, dz}{Q(z)} + \int_{(3)} \frac{P(z) \ln z\, dz}{Q(z)} .$$

Now (1) is on the branch amp $z = 0$ and (3) is on the branch amp $z = 2\pi$, so we have

$$2\pi i\, \mathcal{R} = \int_0^{\infty} \frac{P(x)}{Q(x)} \ln x\, dx + \int_{\infty}^{0} \frac{P(x)(\ln x + 2\pi i)dx}{Q(x)} ,$$

and $$\int_0^{\infty} \frac{P(x)}{Q(x)}\, dx = -\mathcal{R} .$$ Q.E.D.

Appendix C. Explicit Formulas for[1] $f_L(a_r; u, s)$ for $a_r = 0, 1, \ldots 2s+1$, and $s = 0, 1, 2$.

The following formulas are derived from equation (3.8.16).

Case 1. $s = 0$ (Groups of size 1).

$$f_L(0; u, 0) = \frac{1}{1 - u} + \frac{u}{(1 - u)^2} \ln u$$

$$f_L(1; u, 0) = - \frac{u}{1 - u} - \frac{u}{(1 - u)^2} \ln u$$

Case 2. $s = 1$ (Groups of size 3).

$$f_L(0; u, 1) = \frac{u^2}{(1 - u)^3} \left\{ \frac{3(u - 3)}{u(1 - u)} - \frac{60}{(1 - u)^2} + \frac{1}{u^2} + \frac{12}{(1 - u)} \left[\frac{2}{1 - u} - \frac{5}{(1 - u)^2} \right] \ln u \right\}$$

$$f_L(1; u, 1) = \frac{u^2}{(1 - u)^3} \left\{ \frac{54}{1 - u} + \frac{36(2 + 3u)}{(1 - u)^2} + \frac{6}{u} + \frac{18}{(1 - u)} \left[1 - \frac{8}{1 - u} + \frac{10}{(1 - u)^2} \right] \ln u \right\}$$

---

[1]Recall: $u = p/(1 - p)$ for all $0 < p < 1$.

$$f_L(2;\, u,\, 1) = \frac{u^2}{(1-u)^3}\left\{ \frac{45(1-3u)}{1-u} - \frac{36(2+3u^2)}{(1-u)^2} - 33 \right.$$

$$\left. + \frac{18}{(1-u)}\left[-3 + \frac{12}{1-u} - \frac{10}{(1-u)^2}\right] \ln u \right\}$$

$$f_L(3;\, u,\, 1) = \frac{u^2}{(1-u)^3}\left\{ \frac{21(-1+3u^2)}{1-u} + \frac{12(2+3u^3)}{(1-u)^2} + 26u \right.$$

$$\left. + \frac{12}{1-u}\left[3 - \frac{8}{1-u} + \frac{5}{(1-u)^2}\right] \ln u \right\}$$

Case 3. $s = 2$ (Groups of size 5).

$$f_L(0;\, u,\, 2) = \frac{1}{2}\,\frac{u^3}{(1-u)^3}\left\{ \frac{5(u^2-5)}{u^2(1-u)} + \frac{60(2u+5)}{u(1-u)^2} + \frac{1260}{(1-u)^3} \right.$$

$$\left. + \frac{840(4+5u)}{(1-u)^4} + \frac{2}{u^3} + \frac{5040}{(1-u)^3}\left[\frac{1}{4} - \frac{4}{3(1-u)} + \frac{3}{2(1-u)^2}\right] \ln u \right\}$$

$$f_L(1;\, u,\, 2) = \frac{15}{2}\,\frac{u^3}{(1-u)^5}\left\{ \frac{5(u-5)}{u(1-u)} - \frac{770}{(1-u)^2} + \frac{350(3-5u)}{(1-u)^3} \right.$$

$$\left. - \frac{280(4+5u^2)}{(1-u)^4} + \frac{1}{u^2} + \frac{60}{(1-u)^2}\left[2 - \frac{21}{1-u} + \frac{56}{(1-u)^2} - \frac{42}{(1-u)^3}\right] \ln u \right\}$$

$$f_L(2;\, u,\, 2) = \frac{60\, u^3}{(1-u)^5} \left\{ \frac{125}{3(1-u)} + \frac{65(2+5u)}{(1-u)^2} + \frac{245}{2}\,\frac{(-3+5u^2)}{(1-u)^3} \right.$$

$$+ \frac{70(4+5u^3)}{(1-u)^4} + \frac{1}{u} + \frac{5}{(1-u)} \left[1 - \frac{24}{(1-u)} + \frac{126}{(1-u)^2} - \frac{224}{(1-u)^3} \right.$$

$$\left. \left. + \frac{126}{(1-u)^4} \right] \ln u \right\}$$

$$f_L(3;\, u,\, 2) = \frac{300\, u^3}{(1-u)^5} \left\{ \frac{77(1-5u)}{12(1-u)} - \frac{47(2+5u^2)}{2(1-u)^2} + \frac{63(3-5u^3)}{2(1-u)^3} \right.$$

$$- \frac{14(4+5u^4)}{(1-u)^4} - \frac{137}{60} + \frac{1}{(1-u)} \left[-5 + \frac{60}{1-u} - \frac{210}{(1-u)^2} + \frac{280}{(1-u)^3} \right.$$

$$\left. \left. - \frac{126}{(1-u)^4} \right] \ln u \right\}$$

$$f_L(4;\, u,\, 2) = \frac{900\, u^3}{(1-u)^5} \left\{ \frac{19(-1+5u^2)}{8(1-u)} + \frac{37(2+5u^3)}{6(1-u)^2} + \frac{77(-3+5u^4)}{12(1-u)^3} \right.$$

$$+ \frac{7(4+5u^5)}{3(1-u)^4} + \frac{29u}{20}$$

$$\left. + \frac{1}{(1-u)} \left[\frac{5}{2} - \frac{20}{(1-u)} + \frac{105}{2(1-u)^2} - \frac{56}{(1-u)^3} + \frac{21}{(1-u)^4} \right] \ln u \right\}$$

$$f_L(5;\, u,\, 2) = \frac{1260\, u^3}{(1 - u)^5} \left\{ \frac{319(1 - 5u^3)}{504(1 - u)} - \frac{107(2 + 5u^4)}{84(1 - u)^2} + \frac{13(3 - 5u^5)}{12(1 - u)^3} \right.$$

$$- \frac{(4 + 5u^6)}{3(1 - u)^4} - \frac{17\, u^2}{70}$$

$$\left. + \frac{1}{(1 - u)} \left[- \frac{5}{6} + \frac{5}{(1 - u)} - \frac{21}{2(1 - u)^2} + \frac{28}{3(1 - u)^3} - \frac{3}{(1 - u)^4} \right] \ln u \right\}$$

JJG:ccw
8/14/58

ORNL-2629
Physics and Mathematics
TID-4500 (14th ed.)

INTERNAL DISTRIBUTION

1. C. E. Center
2. Biology Library
3. Health Physics Library
4-5. Central Research Library
6. Reactor Experimental Engineering Library
7-26. Laboratory Records Department
27. Laboratory Records, ORNL R.C.
28. A. M. Weinberg
29. L. B. Emlet (K-25)
30. J. P. Murray (Y-12)
31. J. A. Swartout
32. E. H. Taylor
33. E. D. Shipley
34. A. S. Householder
35. C. P. Keim
36. W. H. Jordan
37. S. C. Lind
38. F. L. Culler
39. R. S. Livingston
40. A. H. Snell
41. A. Hollaender
42. M. T. Kelley
43. C. A. Preskitt
44. K. Z. Morgan
45. T. A. Lincoln
46. C. S. Harrill
47. H. E. Seagren
48. C. E. Winters
49. D. Phillips
50. M. L. Nelson
51. J. A. Lane
52. R. A. Charpie
53. M. J. Skinner
54. R. R. Dickison
55. D. A. Gardiner
56. M. A. Kastenbaum
57. A. Downing
58. Ky Fan
59. A. A. Grau
60. G. Atta
61-86. A. W. Kimball
87. D. G. Doherty
88. J. Mendell
89. ORNL - Y-12 Technical Library, Document Reference Section

EXTERNAL DISTRIBUTION

90. Division of Research and Development, AEC, ORO
91-114. Virginia Polytechnic Institute, Blacksburg (1 copy ea. to R. A. Bradley, B. Harshbarger, C. W. Clunies-Ross, H. E. Spencer, R. J. Freund, L. S. Brenna, W. Glenn, H. A. David, and R. Bargmann, 15 copies to Statistics Department)
115. Arizona State College, Tempe (J. E. Freund)
116. University of Florida, Gainesville (W. O. Ash)
117. Assessment Division, Camp Detrick, Frederick, Md. (E. L. Cox)
118. Alfred University, Alfred, N.Y. (H. A. Still)
119-121. National Institute of Health, Bethesda, Md. (1 copy ea. to M. A. Schneiderman, J. Cornfield, and W. Haenszel)
122. Roanoke College, Salem, Va. (R. E. Walpole)
123. Abbott Laboratories, North Chicago, Ill. (E. J. Matson)
124. Louisiana State College, Baton Rouge, La. (W. T. Burnett, Jr.)
125-126. Marquette University, Milwaukee (1 copy ea. to H. P. Pettit and J. Talacko)
127. W. A. Golomski, 26 Sherman Terrace, Madison 4, Wis.
128. H. L. Lucas, Box 5447, Raleigh, N.C.
129-153. J. J. Gart, Dept. of Biostatistics, Johns Hopkins Univ., Baltimore
154-761. Given distribution as shown in TID-4500 (14th ed.) under Physics and Mathematics category (75 copies - OTS)