AN ABSTRACT OF THE THESIS OF

BRADLEY ROY SANDS for the degree DOCTOR OF PHILOSOPHY
(Name) (Degree)

in STATISTICS presented on August 14, 1975
(Major Department) (Date)

Title: ANALYSIS OF BINARY DATA IN THE PRESENCE OF NON-BERNOULLI SOURCES OF VARIATION

Redacted for Privacy

Abstract approved: ____
Dr. Donald A. Pierce

A quantal response model, more general than the usual logistic model, is introduced. This model takes into account sources of variability, or experimental error, other than that arising from variability in response between individual organisms (or other objects on test). It is assumed that this extra source of variation is additive on a logistic scale.

The likelihood function for this model is displayed and procedures relating to a likelihood analysis are discussed. An approximate least squares approach is also introduced. A class of tests and a class of estimators (of the scale parameter for this extra source of variation) are examined.

Analysis of Binary Data in the Presence of Non-Bernoulli Sources of Variation

by

Bradley Roy Sands

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1976

APPROVED:

Redacted for Privacy

Associate Professor of Statistics

in charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented August 14, 1975

Typed by Clover Redfern for Bradley Roy Sands

ACKNOWLEDGMENT

The author would like to gratefully acknowledge his deep debt to Professor Don Pierce for his ever ready guidance and encouragement. The author considers himself very fortunate to have worked with such a gifted and sensitive scholar.

To his wife, Linda, who never suffered in silence, the author offers his heartfelt thanks.

The author would also like to express his appreciation to the many others in the department who gave unhesitating assistance.

Lastly the author acknowledges the financial support provided by a Department of Health, Education and Welfare Biometry Traineeship during these studies.

TABLE OF CONTENTS

LIST OF TABLES

ANALYSIS OF BINARY DATA IN THE PRESENCE OF NON-BERNOULLI SOURCES OF VARIATION

I. INTRODUCTION

In recent years the effects of pollutants in our local and general environment have received greater attention. A natural way to attempt measurement of the toxicity of various pollutants is to adopt and adapt the bioassay techniques (see Finney (1964)) used to measure the biological potency or activity of various treatments. These techniques have been very carefully developed and refined over the years, and have proven to be quite accurate and reliable, especially for pharmaceutical purposes.

However, they do depend to a certain extent on working in a laboratory where uniform conditions are carefully maintained, with highly standardized experimental units. The question asked (and answered) is how does a given treatment compare in potency or activity with some other standard treatment of the same type. Thus, although the absolute potency or activity may vary from lab to lab and time to time, this comparison with the standard treatment will remain unchanged.

The environmental scientist however, is concerned with the absolute potency or activity of a treatment and how this potency depends on the physical environment, the size and sex of the organism, the stage in the life cycle of the organism, the immediate or general history of the organism, etc. Thus concern is not only about treatment level but also about how any number of covariables alter the effective treatment level. Furthermore, it is often necessary to establish the dose response relation over a much wider range of stress levels than would be attempted in a traditional bioassay.

Bioassay techniques easily generalize to estimate parameters associated with treatments and covariables. However, in a disturbing number of cases, the bioassay assumption that all the variability is due to the natural variability between individual organisms in their response to a given treatment/covariable level (Bernoulli-variability) has been rejected by "goodness of fit" tests.

If we were dealing with a single treatment/covariable level then a standard procedure (see David (1951)) would be to use a Lexis ratio as an inflation factor for the covariance associated with the treatment/covariable parameters. In fact Finney (1971) recommends (with reservation) the use of such a ratio as a "heterogeneity factor" over a range of treatment/covariable levels. He suggests that this will take into account interaction between members of a group of organisms which would make their response as a whole different from another group at the same treatment/covariable level, and also uncontrolled (or unmeasured) factors which are different from one group to another at the same treatment/covariate level.

Dews and Berkson (1952) warn of bioassay situations where the experiments were repeated so that an empirical estimate of the covariance of the treatment/covariable parameters could be made. They found these empirical estimates very much different from the theoretical covariances corresponding to them. They suspected that "errors of dosage" (which remained unmodeled) were the cause of this, and warned against depending very heavily on theoretical covariances until they could be checked out empirically.

Cornfield (1952) mentions the possibility of improving estimates of the treatment/covariable parameters by more flexible modeling of the covariance structure of the Binomial variates, but feels it may not be worth the effort since cases exist where maximum likelihood estimates of variances are not efficient.

This thesis is addressed to modeling sources of variation in addition to Bernoulli variation, that is, sources of variation which affect groups as a whole. In a sense, it is an attempt to put Lexis theory into a regression setting.

II. THE HOMOGENEOUS MODEL

This chapter is devoted to establishing notation, and a brief overview of the model and several estimation procedures used when there is no "extra" variation See Cox (1970) and Finney (1971) for more details.

Model M_1

We consider the situation where we have n groups of organisms or other objects, the ith group being of size n_i, and n stimulus levels λ_i (not necessarily distinct). The ith group is exposed to level λ_i and r_i Bernoulli "successes" are observed. We assume that the probability of a Bernoulli success is fixed, say at p_i, for all the objects in the ith group. $(i = 1, \ldots, n)$.

We will formulate the usual "homogeneous" logistic linear model as follows.

Model M_1: For $i = 1, \ldots, n$

$$r_i \sim \mathrm{Bin}(p_i, n_i)$$

where

$$p(\lambda_i) = 1/(1 + \exp(-\lambda_i))$$

$$\lambda_i(\beta) = \Sigma_{j=0}^{m} a_i^j \beta_j$$

Thus

$$r_i/n_i = p_i + \epsilon_i$$

where

$$\epsilon_i \sim WS(0, p_i q_i/n_i)$$

and

$$q(\lambda_i) = 1 - p_i(\lambda_i) .$$

The notation $X \sim WS(\mu, \sigma^2)$, the WS denoting "weak sense," means that $E[X] = \mu$ and $Var[X] = \sigma^2$. For the ith group the a_i^j can be considered to be treatment/covariable levels and the β_j treatment/covariable parameters, $j = 1, \ldots, m$. We make the convention that $a_i^0 = 1$ for all i. We let $A = \ll a_i^j \gg$ be the $n \times (m+1)$ "design" matrix, $A^j = \langle a_i^{(j)} \rangle$ be the jth column of A, and $A_i = \langle a_{(i)}^j \rangle$ be the ith row of A. Further, let $\beta = \langle \beta_j \rangle$, a column vector and define the function $pq(\lambda)$ by $pq(\lambda) = p(\lambda) \cdot q(\lambda) = p(\lambda)(1 - p(\lambda))$.

Thus we can abbreviate the above to

$$r_i/n_i = 1/(1+ \exp(-A_i\beta)) + \epsilon_i$$

$$\epsilon_i \sim WS(0, p_i q_i/n_i)$$

The choice of the logistic function $p(\lambda) = 1/(1+\exp(-\lambda))$ is not without some controversy. In at least some circumstances however it has clear physiological justification (see Berkson (1951)) and it is extremely attractive mathematically. For instance, the above implies

$\lambda(p) = \log[p/(1-p)]$. But

$$p^r(1-p)^{n-r} = (1-p)^n \exp[\lambda r]$$

so that λ is the "natural" parameter associated with the Bernoulli and Binomial distributuions when they are expressed as members of the exponential family. If we calculate $\partial p/\partial\lambda = pq(\lambda)$ and take note of the expression for Bernoulli variance we see another of these nice features perculiar to the logistic function. In any case it is almost impossible to statistically distinguish $p(\lambda)$ from its most popular competitor, the probit model,

$$\bar{p}(\lambda) = \int_{-\infty}^{\lambda-5} \phi(x)dx = \Phi(\lambda-5),$$

where Φ is the cdf of the standard normal distribution.

We now examine three of the many ways to form an estimate $\hat{\beta}$ of β, given a set of observations $\{r_i, n_i, A_i\}_{i=1}^n$.

<u>Maximum Likelihood</u>

For the homogeneous model the likelihood of β given $\{r_i, n_i, A_i\}_{i=1}^n$ can be expressed

$$L(\beta; N, A; R) \propto \prod_{i=1}^{n} f(r_i; n_i, A_i; \beta)$$

where

$$f(r; n, v'; \beta) = p(v'\beta)^r q(v'\beta)^{n-r}$$

$$N = \text{diag} \ll n_i \gg \quad \text{and} \quad R = < r_i >$$

Since we wish to find a value of β which maximizes this likelihood we can examine the gradient of $\phi_1 = \log L$

$$\partial\phi_1/\partial\beta_j = \Sigma_{i=1}^n a_i^j [r_i - n_i p(A_i \beta)] \tag{2.1}$$

$$j = 1, \ldots, n$$

Setting this system to zero and solving for β will involve some sort of iterative process. It will be shown below that ϕ_1 is strictly convex, and that a solution to these likelihood equations, if one exists, is in fact a maximum likelihood estimator. If no solution to the likelihood equations exists, then a finite maximum likelihood estimate of β does not exist; the maximum likelihood estimates of some of the p_i are then zero or one. Before becoming involved in this process, let us examine two other methods for estimating β.

Usual Least Squares

For the homogeneous model the error sum of squares, weighted so each term has variance one, can be expressed

$$\phi_2 = SS(\beta; N, A; R)$$

$$= \Sigma_{i=1}^{n} [(r_i/n_i) - p(A_i\beta)]^2 / [pq(A_i\beta)/n_i]$$

Again we examine the gradient of ϕ_2. Noting that

$$\partial pq/\partial\lambda = pq(\lambda)[q(\lambda) - p(\lambda)]$$

we see that

$$\partial\phi_2/\partial\beta_j = -2\Sigma_i a_i^j[r_i - n_i p_i]$$

$$- \Sigma_i a_i^j[r_i - n_i p_i]^2[q_i - p_i]/[n_i p_i q_i] \qquad (2.2)$$

where

$$p_i = p(A_i\beta) \quad \text{and} \quad q_i = 1 - p(A_i\beta).$$

Now we take the "usual" approach of ignoring the second term of (2.2), which results from the dependence of the weights $\{n_i/pq(A_i\beta)\}_{i=1}^{n}$ on β. This approach reduces the above equation to

$$\partial\phi_2/\partial\beta_j = -2\Sigma_i a_i^j[r_i - n_i p(A_i\beta)]$$

We observe that since this is just -2 times $\partial\phi_1/\partial\beta_j$ above, the maximum likelihood estimator and the usual least squares estimator will coincide. This is similar to results obtained by Bradley (1973) in exponential families.

True Least Squares

To find a true least squares estimator, we would need a value of β which would reduce the expression (2.2) to zero. Rather than persue a true least squares estimator at this point, we simply point out that at the usual least squares estimator (or the maximum likelihood estimator) $\hat{\beta}$ the true gradient (expression 2.2) is not zero but rather

$$-\Sigma_{i=1}^{n} a_i^j \{[(r_i/n_i)-p_i]^2/[p_i q_i/n_i]\}[q_i-p_i]$$

where

$$p_i = p(A_i\hat{\beta}) \quad \text{and} \quad q_i = 1 - p(A_i\hat{\beta}) .$$

This is another weighted error sum of squares where the weights are now $\{n_i a_i^j [q_i-p_i]/p_i q_i\}$ rather than $\{n_i/p_i q_i\}$. Minimizing ϕ_2 tends to minimize this sum of squares also, but that does not reduce the "true" gradient to zero. The last section of Clutton-Brock (1967) discusses the difference between true least squares estimators and the usual least squares estimator for a similar problem.

Estimation

We proceed to describe a procedure to find a value of β which brings $<\partial\phi_1/\partial\beta_j>$ to zero by making use of the Taylor's series approximation

$$p(A_i\beta) \doteq p(A_i\beta^*) + pq(A_i\beta^*)A_i(\beta-\beta^*)$$

Given a guess β^* we solve the system

$$\Sigma_{i=1}^{n} a_i^j[r_i - n_i p(A_i\beta^*)] = \Sigma_{i=1}^{n} a_i^j n_i pq(A_i\beta^*)A_i(\beta-\beta^*)$$

for $\beta - \beta^*$ and hence for β. This system can be expressed more compactly

$$A'Y^* = A'ND^*A(\beta-\beta^*)$$

so

$$\beta = [A'ND^*A]^{-1}A'Y^* + \beta^*$$

where

$$Y^* = \langle r_i - n_i p(A_i\beta^*)\rangle$$

$$N = \text{diag} \ll n_i \gg$$

and

$$D^* = \text{diag} \ll pq(A_i\beta^*) \gg$$

Now, since we introduced the Taylor's approximation, β is not the solution desired but simply a better guess than β^* was. We let β play the role that β^* did and repeat the above until $\beta - \beta^*$ gets small.

We can view the above as iterated weighted regression of $(ND^*)^{-1}Y^*$ onto $\underline{R}(A)$ or iterated standard regression of $(ND^*)^{-1/2}Y^*$ onto $\underline{R}(X)$, where $X = (ND^*)^{1/2}A$ and $\underline{R}(X)$ denotes the range of the column space of the matrix X.

Alternatively we can view the above process as using a second order Taylor's series approximation to ϕ_1 at β^* so that β will maximize this quadratic. Since ϕ_1 is not quadratic in β we must again repeat (iterate) this process to reach the estimator $\hat{\beta}$.

After $\beta - \beta^*$ has become small, we infer that

$$\Sigma a_i^j[r_i - n_i p(A_i\beta^*) - n_i pq(A_i\beta^*)A_i(\beta-\beta^*)]$$

$$\doteq \Sigma a_i^j[r_i - n_i p(A_i\beta)]$$

is a good approximation and that the most recent guess for β will cause the latter expression to be very close to zero, for all j. We denote by $\hat{\beta}$ this most recent guess of β, and estimate $\text{Cov}(\hat{\beta})$ by $(A'N\hat{D}A)^{-1}$. This is appropriate from the maximum likelihood viewpoint since $A'N\hat{D}A$ is the Hessian of ϕ_1, which in this case does not involve R and is hence Fisher's information matrix, evaluated at $\beta = \hat{\beta}$. It is also appropriate from the weighted least squares viewpoint, since $\text{Var}[(ND^*)^{-1/2}Y^*] \doteq I$ and $(X'X)^{-1} = AND^*A$, where $X = (ND^*)^{1/2}A$. These procedures depend heavily upon the modeling assumptions regarding the covariance structure of the $\{r_i\}$ however. Wisdom generated by past disasters impels us to try to check the plausibility of $\text{Cov}(\hat{\beta}) = (A'N\hat{D}A)^{-1}$.

Note that since $A'ND^*A$ is always positive definite (provided A is of full rank) ϕ_1 is a concave function (and ϕ_2 convex). Thus when we find a zero of the gradient by the above process we know that we have a maximum (or minimum), and in fact the only such. This fact, and the variety of ways to find reasonable starting points, have undoubtedly contributed to the success this procedure has enjoyed in practice.

Goodness of Fit

A simple test of our assumptions is the usual chi-squared test using $\phi_2(\hat{\beta})$. Two drawbacks of this test are that its power is not directed toward any special alternatives, and that if the values of $p(A_i\hat{\beta})$ approach zero or one to closely the binomial distribution of the residuals becomes quite skewed, so that unless the n_i are very large, the normal approximations involved in using chi-squared critical values become problematical.

If, as a chi-squared variate with $n - (m+1)$ degrees of freedom, $\phi_2(\hat{\beta})$ is significantly large, then we suspect either a random or systematic deviation from our assumptions.

A random deviation generally is associated with one of several causes. Perhaps objects in a group have interacted in some way, so that the group as a whole has a different underlying probability of "success." More generally this probability may be altered by any number of unmeasured covariables that are constant for all objects in a group but vary from group to group. The increased dispersion of r_i/n_i about $p(A_i\hat{\beta})$ that these uncontrolled factors cause inflates the value of $\phi_2(\hat{\beta})$. For a fixed $p(A_i\hat{\beta})$ (that is, fixed A_i) this increased dispersion is termed Lexis variation, and could be compensated for by an inflation factor called the Lexis ratio calculated from $\phi_2(\hat{\beta})$. The Lexis ratio would be used to inflate all the variances involved (see David (1951)). Finney (1970) implicitely assumes that the $p(A_i\hat{\beta})$ do not vary enough to discredit the use of such an inflation factor, which he calls a "heterogeneity factor" $h = \phi_2(\hat{\beta})/[n-(m+1)]$.

A systematic deviation is generally caused by either the need for some transformation of one or several of the treatment/covariable levels A^j, or the need for some other stimulus-response relation $p(\lambda)$.

Random and systematic deviations are difficult to differentiate. The penalty for improperly ascribing a large $\phi_2(\hat{\beta})$ to a random deviation ranges from not finding a transformation which would lead to a better understanding of the relationships involved, to a very poor extrapolation which grossly mis-estimates the expected response at certain treatment levels. The decision must be based on all information available, including graphs, expert opinion and the history of analyses of the same "sort" of data.

The rest of this thesis is devoted to examining ways to handle random deviations with more care.

III. THE HETEROGENEOUS MODEL

Here we introduce a more general model, which incorporates an extra source of variation, in an attempt to get a better grasp on the possible structure of $\mathrm{Cov}(\hat{\beta})$. There has been remarkably little in the literature on attempts to model such extra variation. Patwary and Haley (1967) have developed a model which has similarities to that given here, but consists of a special Poisson-type source of variation. The plausibility of our modeling of this extra source of variation will need to be critically examined for each class of applications before it is used routinely, but the choice seems to be as reasonable as the logistic linear model itself.

We also discuss the likelihood function and its differentials, reconsider the goodness of fit test and discuss special aspects of an algorithm for maximizing the likelihood.

Model M_2

We formulate the heterogeneous logistic linear model as follows.

Model M_2: For $i = 1, \ldots, n$ let

$$\lambda_i(\beta, \sigma) = A_i\beta + \sigma e_i$$

where the e_i are independent random variables, $e_i \sim WS(0, 1)$, and σ is a new parameter. Conditionally upon the e_i, let the r_i

be independent,

$$r_i | e_i \sim \text{Bin}(p_i, n_i)$$

where

$$p_i = p(\lambda_i) = 1/(1+\exp(-\lambda_i))$$

$$i = 1, 2, \ldots, n.$$

Thus when $\sigma = 0$ this is the model M_1; when $\sigma > 0$ the model has an extra source of variation which is homogeneous and additive on the logistic scale.

It is worthwhile to note that a Taylor's approximation gives

$$\text{Var}[r_i/n_i] \doteq \frac{p_i q_i}{n_i} (1+\sigma^2 n_i p_i q_i)$$

This will be elaborated upon in Chapter IV.

<u>Likelihood Function</u>

We now define $\rho = \sigma^2$ and develop the likelihood function for β and ρ.

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

$$h_i(r_i, e_i; \beta, \rho) = p_i^{r_i} q_i^{n_i - r_i}$$

where

$$p_i(e_i; \beta, \rho) = 1/(1+\exp(-A_i\beta-\sigma e_i)$$

and

$$q_i(e_i; \beta, \rho) = 1 - p_i(e_i; \beta, \rho)$$

However, not having been able to observe e_i, we base our inference on

$$f_i(r_i; \beta, \rho) = \int h_i(r_i, e; \beta, \rho) dF(e)$$

where F is a specified distribution function for e.

Thus the joint likelihood will be

$$L(\beta, \rho | R) = \prod_{i=1}^{n} f_i(r_i; \beta, \rho)$$

where $R = \langle r_i \rangle$. Finally we define

$$\phi = \log L(\beta, \rho | R).$$

Differential Structure of $\phi = \log L$

Before displaying the gradient and Hessian of ϕ, we consider some preliminary calculations and definitions. We have that

$$\partial p / \partial \lambda = pq(\lambda)$$

$$\partial pq / \partial \lambda = pq(\lambda)[q(\lambda) - p(\lambda)]$$

so

$$\partial h / \partial \lambda = h[r - np]$$

and

$$\partial^2 h/(\partial\lambda)^2 = h[(r-np)^2-npq]$$

Note that these results are of peculiarly simple form due to our choice of the logistic stimulus-response relationship.

Now define

$$\theta = (\theta_1, \ldots, \theta_{m+2})$$
$$= (\rho, \beta_0, \ldots, \beta_m)$$

so that

$$\partial\lambda_i/\partial\theta_1 = e/2\theta_1^{1/2}$$

and

$$\partial\lambda_i/\partial\theta_s = a_i^{s-2} \quad \text{for} \quad s = 2, \ldots, m+2.$$

Further, we define

$$f^{(s)} = \partial f/\partial\theta_s$$
$$= \int h[r-np][\partial\lambda/\partial\theta_s]dF(e)$$
$$f^{(s,t)} = \partial^2 f/\partial\theta_s\partial\theta_t$$
$$= \int h[(r-np)^2-npq][\partial\lambda/\partial\theta_s][\partial\lambda/\partial\theta_t]dF(e)$$

Now, we can write

$$\partial\phi/\partial\theta_s = \Sigma_{i=1}^n f_i^{(s)}/f_i$$

and

$$\partial^2\phi/\partial\theta_s\partial\theta_t = \Sigma_{i=1}^n [f_i f_i^{(s,t)} - f_i^{(s)} f_i^{(t)}]/f_i^2$$

Finally we define

$$g_n(\theta; R) = \langle \partial\phi/\partial\theta_s \rangle_{s=1}^{m+2}$$

$$G_n(\theta; R) = \langle\langle \partial^2\phi/\partial\theta_s\partial\theta_t \rangle\rangle_{s,t=1}^{m+2}$$

Note that since G is not always negative definite, ϕ is not always concave and we may mistake inflection points for maximums. This is a difficulty both in assuring that a solution of the likelihood equation is a maximum likelihood estimate, and in algorithms for finding a solution to the likelihood equation. The latter point is dealt with below. For the former we can only rely on the hope that for moderate amounts of data the likelihood function will be unimodal in a substantial neighborhood of the maximum likelihood estimate. This hope is not unfounded, of course, but is based on conventional asymptotic results which are considered more in Chapter V.

For our analysis it seems that the most plausible choice for the distribution of e would be $e \sim N(0, 1)$. Under this choice we cannot evaluate the integrals in ϕ, $f^{(s)}$ or $f^{(s,t)}$ analytically but we can make use of a numerical integration technique. Since we can use Hermitian quadrature (see Abramowitz and Stegun (1966)) to make the approximation, for a given function f,

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2)dx \doteq \Sigma_t H_t f(x_t)$$

where H_t and x_t are tabulated constants, it follows that

$$\int_{-\infty}^{\infty} f(e)(2\pi)^{-1/2} \exp(-e^2/2)de \doteq \Sigma_t [(2\pi)^{-1/2} H_t] f(2^{1/2} x_t) .$$

We have found that this approximation can be used very efficiently, even when using as many as 20 terms in the above summation.

Goodness of Fit Reconsidered

With this new model and likelihood function in hand, we are in a position to reconsider our goodness of fit test.

In a sense the usual chi-squared goodness of fit test is usually thought of as testing model M_1 against "any other model." Tests which pit M_1 against M_2 should be somewhat more sensitive since the alternatives are so much more restricted. These tests will take the form $\rho = 0$ vs. $\rho > 0$, treating β as a nuisance parameter.

After the M_1 maximum likelihood estimate (m.l.e.) $\hat{\beta}^0$ and the M_2 m.l.e. $\hat{\beta}$, $\hat{\rho}$ have been found, we can form a likelihood ratio test, or simply examine the marginal distribution of $\hat{\rho}$

and test $\hat{\rho} = 0$. However we can consider at least one test which requires only the M_1 m.l.e. $\hat{\beta}^0$. This is considerably more attractive than the above since the M_2 likelihood not only involves quadratures, but also is not concave, and hence is much more difficult to maximize than the M_1 likelihood.

We consider a special case of Neyman's locally asymptotically optimal test of class $C(\alpha)$, (see Moran (1970)). The test statistic is of the form

$$T = \partial\phi/\partial\rho|_{\hat{\beta},0} - \Sigma_{j=0}^{m} b_j \partial\phi/\partial\beta_j|_{\hat{\beta},0}$$

where $\hat{\beta}$ is a root-n consistent estimator of β, and the $\mathbf{b}_j$ are chosen to minimize $\mathrm{Var}[T]$. Since the M_1 m.l.e. $\hat{\beta}^0$ is root-n consistent under the null hypothesis, and

$$\partial\phi/\partial\beta_j|_{\hat{\beta}^0,0} = 0, \qquad j = 0,\ldots,m$$

such a choice simplifies the above expression to

$$\begin{aligned} T &= \partial\phi/\partial\rho|_{\hat{\beta}^0,0} \\ &= \Sigma_{i=1}^{n} f_i^{(1)}/f_i|_{\hat{\beta}^0,0} \\ &= \Sigma_{i=1}^{n} \left.\frac{\int h_i[r_i - n_i p_i]\, e\, dF(e)}{2\rho^{1/2}\int h_i\, dF(e)}\right|_{\hat{\beta}^0,0} \end{aligned}$$

Now since $E(e) = 0$ we cannot simply substitute $\rho = 0$ in numerator and denominator without obtaining the indeterminant form $0/0$. L'Hospital's rule can be used as follows.

We suppose that $\hat{\beta}^0$ has been substituted into the numerator, so that it is a function of ρ. Then the derivative of the numerator will be

$$\frac{1}{2}\int h_i[(r_i - n_ip_i)^2 - n_ip_iq_i][e^2/\rho^{1/2}]dF(e)$$

since $\partial h[r-np]/\partial\lambda = h[(r-np)^2 - npq]$.

The derivative of the denominator will be

$$\int h_i dF(e)/\rho^{1/2} + \int h_i[r_i - n_ip_i]edF(e)$$

so that the quotient becomes

$$\frac{\int h_i[(r_i - n_ip_i)^2 - n_ip_iq_i]e^2 dF(e)}{2\int h_i dF(e) + 2\rho^{1/2}\int h_i[r_i - n_ip_i]edF(e)}$$

which approaches (as $\rho \to 0$)

$$[(r_i - n_i\hat{p}_i)^2 - n_i\hat{p}_i\hat{q}_i]/2$$

where now $\hat{p}_i = p(A_i\hat{\beta}^0)$.

Thus we have

$$T = \frac{1}{2}\Sigma_{i=1}^{n} [(r_i - n_i\hat{p}_i)^2 - n_i\hat{p}_i\hat{q}_i]$$

When T is large (positive) then we feel that the likelihood could be significantly improved at values of $\hat{\rho}$ greater than zero, so $\rho > 0$. However when T is small (negative) we feel that the likelihood could not be improved at values of $\hat{\rho}$ greater than zero, hence $\rho = 0$. The problem now facing us is finding a critical value for the statistic T.

One way is to approximate the moments of T when $\rho = 0$ by substituting p_i for $\hat{p}_i$, so that we have (for $i \neq j$)

$$\text{Cov}[(r_i - n_i p_i), (r_j - n_j p_j)] = 0.$$

Then $E[T] = 0$ and

$$\text{Var}[T] = \frac{1}{2}\Sigma_{i=1}^{n} [2(n_i p_i q_i)^2 - n_i p_i q_i (1 - 6p_i q_i)] .$$

This follows since

$$E[(x^2 - E(x^2))^2] = E(x^4) - E^2(x^2)$$

$$E[(r-np)^2] = 3(npq)^2 + npq(1-6pq)$$

and

$$E^2[(r-np)^2] = (npq)^2$$

so

$$E[((r-np)^2-npq)^2] = 3(npq)^2 + npq(1-6pq) - (npq)^2$$
$$= 2(npq)^2 + npq(1-6pq).$$

These first two moments can be used for the standard practice of fitting a scaled chi-squared distribution and establishing a critical value for T.

Another way uses the approximate normality of the variates

$$Z_i = (r_i - n_i\hat{p}_i)/(n_i\hat{p}_i\hat{q}_i)^{1/2}$$

and takes into account their correlation. This will develop from the analysis in Chapter IV.

The Newton-Raphson Algorithm

In searching for a value of θ which maximizes the log likelihood function ϕ, suppose we have a first guess (or starting value θ^*. We want to find a vector $\theta - \theta^*$ which will step us to an improved guess θ. Normally we would use

$$\phi(\theta) \doteq \phi(\theta^*) + (\theta-\theta^*)'g(\theta^*) + \frac{1}{2}(\theta-\theta^*)'G(\theta^*)(\theta-\theta^*).$$

Taking the gradient we obtain

$$g(\theta) \doteq g(\theta^*) + G(\theta^*)(\theta-\theta^*).$$

Setting $g(\theta) = 0$ we find

$$\theta - \theta^* = [-G(\theta^*)]^{-1} g(\theta^*).$$

However, the Hessian $G(\theta)$ is not always negative definite. If θ is a value which reduces the gradient of a quadratic to zero, then θ may maximize this quadratic or it may simply be a saddle point of the quadratic.

Nonetheless, iterative use of quadratic approximations gives very fast convergence (when it does converge at all) so we feel that rather than resort to lower order search methods, we will alter the Newton-Raphson algorithm to make it more robust.

We break the search for $\theta - \theta^*$ into two parts. First we use a modified quadratic approximation at θ^* to determine a search direction d and a first guess about the step length α_0. We then refine α_0 by cubic interpolation to find the step length α such that $\theta - \theta^* = \alpha d$ steps us to an improved value of ϕ.

The Search Direction d

We will define the search direction d by

$$d = [\bar{G}(\theta^*)]^{-1} g(\theta^*) = [-G(\theta^*) + D(\epsilon)]^{-1} g(\theta^*)$$

The matrix $D(\epsilon)$ is chosen to make $\bar{G}$ positive definite and also so that d will be roughly a ridge estimator (see Marquardt (1970)).

The procedure we have used to find $D(\epsilon)$, or rather $\overline{G}(\theta^*)$ is based on the following.

Suppose $G - \epsilon I$ was negative definite. Then we could use a Cholesky decomposition to obtain $-G + \epsilon I = TT'$ where T is a lower triangular matrix and T' is its transpose. Thus for each $i = 1, \ldots, m+2$

$$t_{ii} = (-g_{ii} + \epsilon - \Sigma_{j=1}^{i-1} t_{ij}^2)^{1/2}$$

and

$$t_{ki} = (-g_{ki} + \epsilon - \Sigma_{j=1}^{i-1} t_{kj} t_{ij}) / t_{ii}$$

for $k = i+1, \ldots, m+2$.

Now we modify the above in case $G - \epsilon I$ is not "negative definite enough." Suppose that the first $i-1$ columns of T have been calculated, and that $|t_{ij}| \leq b_1$ for $i = 2, \ldots, m+2$ and $j = 1, \ldots, i-1$, b_1 being chosen so that if $G - \epsilon I$ is sufficiently negative definite then $TT' = -G + \epsilon I$ and so that upper bounds on the diagonal elements of T are minimized (see Murray (1972)). We do this by setting

$$b_1^2 = \max\{\max_i\{|g_{ii}|\}, \frac{1}{m+2} \max_{i \neq j}\{|g_{ij}|\}\}$$

We define

$$\overline{t}_i = \max\{|-g_{ii} + \epsilon - \Sigma_{j=1}^{i-1} t_{ij}^2|^{1/2}, 2^{-16}, \epsilon$$

and

$$\bar{t}_k = (-g_{ki} + \epsilon - \Sigma_{j=1}^{i-1} t_{kj} t_{ij}) / \bar{t}_i$$

for $k = i+1, \ldots, m+2$. Let

$$\bar{b}_3 = \max\{|\bar{t}_k| : k = i+1, \ldots, m+2\}$$

If $\bar{b}_3 \leq b_1$ then

$$t_{ki} = \bar{t}_k \quad \text{for} \quad k = i, \ldots, m+2$$

If $\bar{b}_3 > b_1$ then

$$t_{ii} = \bar{b}_3 \bar{t}_i / b_1$$

and

$$t_{ki} = b_1 \bar{t}_k / \bar{b}_3 \quad \text{for} \quad k = i+1, \ldots, m+2$$

In either case

$$|t_{ki}| \leq b_1, \quad k = i+1, \ldots, m+2$$

Note that even when $\bar{b}_3 > b_1$ these changes affect only the diagonal elements of TT'.

This modification of the Cholesky decomposition is essentially that used in Murray's (1972) "numerically stable modified Newton method based on Cholesky factorization." We simply set $\bar{G}(\theta^*) = TT'$ so that

$$d = (TT')^{-1} g(\theta^*).$$

Where the algorithm in Chapter II could be considered to be among the class of steepest descent algorithms with norm $\|d\| = d'G(\beta)d$, the above modification can be considered to be a steepest descent algorithm with respect to the norm $\|d\| = d'[G+D]d = d'TT'd$.

A program for carrying out this decomposition and inversion, $(TT')^{-1}$, is included in the Appendix.

<u>The Step Length α</u>

Once a direction d has been chosen a basic step length λ is chosen, usually one to begin. ϕ and $d'g$ are evaluated at θ^* and $\theta = \theta^* + \lambda d$. We denote these values ϕ^*, $d'g^*$, ϕ and $d'g$ respectively.

If $\phi - \phi^* > 0$ and $d'g > 0$ we leave λ the same, replace θ by $\theta^* + 2\lambda d$ and θ^* by $\theta^* + \lambda d$, evaluate ϕ and $d'g$ at θ^* and θ, as above, and test again.

If $\phi - \phi^* \leq 0$ and $d'g > 0$ we replace θ by $\theta^* + .2\lambda d$ and λ by $.2\lambda$, evaluate ϕ and $d'g$ again and test again.

If $d'g < 0$ we perform a cubic interpolation. The slope of the cubic function which passes through ϕ^* and ϕ with slopes $d'g^*$ and $d'g$ respectively can be expressed (see Davidon (1959))

$$m(\alpha) = d'g^* - 2\alpha(d'g^*+Z)/\lambda + \alpha^2(d'(g^*+g)+2Z)/\lambda^2$$

where

$$Z = 3(\phi^*-\phi)/\lambda + d'(g^*+g)$$

As α runs from 0 to λ, θ runs from θ^* to $\theta^* + \lambda d$ along $\theta^* + \alpha d$.

This slope is zero when $\alpha = \lambda(1-a)$ where

$$a = (d'g+Q-Z)/(d'(g-g^*)+2Q)$$

and

$$Q = -(Z^2-(d'g^*)(d'g))^{1/2} .$$

A program for maximizing ϕ over β, ρ (when $\rho > 0$), incorporating the above modifications, is included in the Appendix.

IV. LEAST SQUARES

In this chapter we investigate sources of information about σ^2 (and hence $\mathrm{Cov}(\hat{\beta})$) which are available without resorting to the full likelihood analysis described in the previous chapter. We assume that we have an (iterated weighted) least squares estimate $\hat{\beta}^0$, under the constraint $\sigma^2 = 0$ (recall that this is the same as a maximum likelihood estimate when $\sigma^2 = 0$), and approximate our model, in a neighborhood of $\beta = \hat{\beta}^0$, $\sigma^2 = 0$, by a linear model.

In this framework we use "residuals" for two purposes. First we compare a variety of tests of $\sigma^2 = 0$ vs. $\sigma^2 > 0$. Secondly we compare estimators of σ^2. A good (not necessarily best) estimate of σ^2 will allow us to revise our weights and perform another iteration or two for a better (iterated weighted) least squares estimate $\hat{\beta}^+$ and, most important, obtain a more reliable estimate of $\mathrm{Cov}(\hat{\beta})$.

The Linear Approximation

We suppose that we have an (iterated weighted) least squares estimator $\hat{\beta}^0$ under the constraint $\sigma^2 = 0$. Let $\lambda_i^0 = A_i\hat{\beta}^0$.

Recall model M_2: For $i = 1, \ldots, n$ let

$$\lambda_i(\beta, \sigma) = A_i\beta + \sigma e_i, \quad e_i \sim WS(0, 1)$$

and

$$r_i | e_i \sim \text{Bin}(p_i, n_i)$$

so

$$r_i/n_i = p(A_i\beta+\sigma e_i) + \epsilon_i$$

If we consider for the moment that β and σ are the "true" parameters, then a Taylor's series approximation gives us

$$p(A_i\beta+\sigma e_i) \doteq p(A_i\beta) + pq(A_i\beta)\sigma e_i \ ,$$

and the variance of p due to e can be approximated

$$V[p] \doteq (\sigma pq)^2$$

Thus

$$V[r_i/n_i] \doteq V[p(A_i\beta+\sigma e_i)] + V[\epsilon_i]$$

$$\doteq \sigma^2 p_i^2 q_i^2 + p_i q_i/n_i$$

$$\doteq (p_i q_i/n_i)[\sigma^2 n_i p_i q_i + 1]$$

where

$$p_i = p(A_i\beta) \ .$$

More generally we have

$$p(A_i\beta+\sigma e_i) \doteq p(\lambda_i^0) + pq(\lambda_i^0)[A_i(\beta-\hat{\beta}^0)+\sigma e_i]$$

for small $\beta - \hat{\beta}^0$ and small σ.

Thus

$$r_i/n_i \doteq p(\lambda_i^0) + pq(\lambda_i^0)[A_i(\beta - \hat{\beta}^0) + \sigma e_i] + \epsilon_i$$

or equivalently, weighting for the Bernoulli variance,

$$\frac{r_i/n_i - p_i^0}{(pq_i^0/n_i)^{1/2}} \doteq (n_i pq_i^0)^{1/2} A_i(\beta - \hat{\beta}^0)$$

$$+ \sigma(n_i pq_i^0)^{1/2} e_i + (n_i/pq_i^0)\epsilon_i$$

where

$$p_i^0 = p(\lambda_i^0) = p(A_i\hat{\beta}^0)$$

$$pq_i^0 = pq(\lambda_i^0) = p(\lambda_i^0)[1 - p(\lambda_i^0)]$$

Define

$$Y_i = [(r_i/n_i) - p(\lambda_i^0)]/[pq(\lambda_i^0)/n_i]^{1/2}$$

$$N = \text{diag} \ll n_i \gg$$

$$\mathbf{D}^0 = \text{diag} \ll pq(\mathbf{A}_i\hat{\beta}^0) \gg$$

$$\mathbf{X}^0 = [N\mathbf{D}^0]^{1/2}\mathbf{A}$$

$$Y = <Y_i>, \quad e = <e_i>, \quad \epsilon = <\epsilon_i> .$$

Now we can express the above more compactly as

$$Y \doteq \mathbf{X}^0(\beta - \hat{\beta}^0) + \sigma[N\mathbf{D}^0]^{1/2} e + N^{1/2}[\mathbf{D}^0]^{-1/2}\epsilon . \qquad (4.1)$$

Note that

$$\mathrm{Cov}[[ND^0]^{1/2}e] = ND^0$$

and

$$\mathrm{Cov}[N^{1/2}[D^0]^{-1/2}\epsilon] = I.$$

We will generally ignore the dependence of X^0 and D^0 on $\hat{\beta}^0$.

In a model of the form $Y \sim WS(X\beta, \sigma^2 ND+I)$ there are various justifications for basing inference about σ^2 on the simple least squares residuals QY, $Q = I - X(X'X)^{-1}X'$. Certainly, a linear function of Y whose expectation is not a function of β must be a function of QY. Seely (1972) shows that if Y is multivariate normal then QY is a maximal invariant, where the group of transformations is the set of affine linear transformations on β. We shall consider methods of inference about σ^2 which are based on quadratic forms in QY.

In (4.1) $QY = Y$ since the maximum likelihood equations can be written as (2.1)

$$Y'[ND]^{1/2}A = Y'X^0 = \phi$$

under the null hypothesis $\sigma^2 = 0$. Thus we have approximately that

$$Y = QY \sim WS(\phi, \sigma^2 QNDQ+Q)$$

Since $\sigma^2 QNDQ + Q$ is of rank $n-(m+1) < n$ we further reduce our statistic by mapping the "error" space $\subset \mathbb{R}^n$ to $\mathbb{R}^{n-(m+1)}$ such that the coordinates of the image in $\mathbb{R}^{n-(m+1)}$ are uncorrelated.

To do this we construct a matrix C whose columns C^i are the "non-trivial" characteristic vectors of $QNDQ$. Let Γ denote the diagonal matrix of corresponding characteristic values, $\text{diag}\ll\gamma_i\gg$. It follows that $C': \mathbb{R}^n \to \mathbb{R}^{n-(m+1)}$ such that

i) C is $n \times (n-(m+1))$

ii) $QNDQC = C\Gamma$

iii) $C^i \in \underline{R}(Q)$

iv) $QC = C$

v) $C'QC = I$

vi) $C'QNDQC = \Gamma$

vii) $CC' = Q$

viii) $C\Gamma C' = QNDQ$.

The "observations" which we now wish to analyze are $\mathbf{Z}$, where

$$\mathbf{Z} = C'QY = C'Y \sim WS(\phi, \sigma^2\Gamma+I)$$

Note that $\sigma^2\Gamma + I$ is diagonal and of full rank.

In comparing the entries of ND and of $C'NDC = \Gamma$ it is worth noting a theorem of Anderson (1971) that if $\gamma_1, \ldots, \gamma_{n-(m+1)}$ are the elements of Γ, in increasing order, and if $\eta_1, \eta_2, \ldots, \eta_n$ are the elements of ND, in increasing order, then

$$\eta_i \le \gamma_i \le \eta_{i+(m+1)}, \quad \text{for} \quad i = 1, \ldots, n-(m+1) .$$

In this sense the set $\{\gamma_i\}$ "resemble" the set $\{\eta_i\}$.

The Testing Problem

Let us examine the problem of testing $\sigma^2 = 0$ vs. $\sigma^2 > 0$ when we have observations

$$Z \sim N(\phi, \sigma^2\Gamma + I).$$

The likelihood ratio statistic for testing $\sigma^2 = 0$ vs. $\sigma^2 = \sigma^2_A$ is

$$S(Z, \sigma^2_A) = \Sigma_{i=1}^n \frac{\gamma_i Z_i^2}{1+\sigma^2_A \gamma_i}$$

Since this depends on σ^2_A we cannot have a uniformly most powerful test of $\sigma^2 = 0$ vs. $\sigma^2 > 0$. Lacking such a test, the β-optimal, level α tests as described by Davies (1969) form a class of attractive alternatives.

Definition: For an arbitrary α, β with $0 \leq \alpha < \beta \leq 1$, a test of the hypothesis $\theta = 0$ vs. $\theta > 0$ will be called β-optimal, level α if its power function $B(\theta)$ has the following properties:

i) $B(0) \leq \alpha$, and

ii) $\inf\{\theta_0 : \theta \geq \theta_0 \Rightarrow B(\theta) \geq \beta\}$

is minimized among all tests whose power satisfies i).

Let

$$B(\sigma^2_A, c; \sigma^2) = \Pr\{S(Z, \sigma^2_A) > c \mid Z \sim N(\phi, \sigma^2\Gamma+I)\} \qquad (4.2)$$

In particular

$$B(\sigma_A^2, c; 0) = \Pr\{\Sigma \frac{\gamma_i U_i^2}{1+\sigma_A^2 \gamma_i} > c \mid U \sim N(\phi, I)\}$$

and

$$B(\sigma_A^2, c; \sigma_A^2) = \Pr\{\Sigma\ \gamma_i U_i^2 > c \mid U \sim N(\phi, I)\}$$

Thus given a value for β we can choose c such that $B(\sigma_A^2, c; \sigma_A^2) = \beta$, and then choose σ_A^2 such that $B(\sigma_A^2, c; 0) = \alpha$. We have associated with each pair α, β one and only one pair c, σ_A^2 such that $B(\sigma_A^2, c; \sigma_A^2) = \beta$ and $B(\sigma_A^2, c; 0) = \alpha$. After we note that $B(\sigma_A^2, c; \sigma^2)$ is a monotone non-decreasing function of σ^2 for each $\sigma_A^2 > 0$, we are assured by a theorem of Davies ((1969), pg 526) that the likelihood ratio $S(Z, \sigma_A^2)$ provides a β-optimal, level α test of $\sigma^2 = 0$ vs. $\sigma^2 > 0$. Thus a likelihood ratio test reaches a given power level "first" (that is with smallest true σ^2) among all level α tests.

Various insights are provided regarding these tests by expressing them in terms of

i) $U \sim N(\phi, I)$

ii) $Z \sim N(\phi, \sigma^2\Gamma + I)$

iii) $Y = \langle (r_i/n_i - p_i)/(p_i q_i/n_i)^{1/2} \rangle$

"U" Critical Regions

In terms of $U \sim N(\phi, I)$ our hypotheses can be expressed $H_0: Z = U$ vs. $H_A: Z = \sqrt{(\sigma^2\Gamma+I)}U, (\sigma^2 > 0)$. The β-optimal, level α critical regions can be expressed

$$R(\sigma_A^2) = \{U'\Gamma(\sigma_A^2\Gamma+I)^{-1}(\sigma^2\Gamma+I)U > c\} .$$

This expression is convenient when searching for values of σ_A^2 and c to "go with" given α and β values.

"Z" Critical Regions

In terms of $Z \sim N(\phi, \sigma^2\Gamma+I)$ our hypotheses are simply $H_0: \sigma^2 = 0$ vs. $H_A: \sigma^2 > 0$, and all β-optimal, level α critical regions can be expressed

$$R(\sigma_A^2) = \{Z'\Gamma(\sigma_A^2\Gamma+I)^{-1}Z > c\} .$$

Note that the efficient score in this context is

$$\partial \ln f/\partial\sigma^2 = \frac{1}{2} \Sigma_{i=1}^{n} (\sigma_A^2\gamma_i+1)^{-2}\gamma_i Z_i^2 - \mathrm{tr}[(\sigma_A^2\Gamma+I)^{-1}\Gamma]$$

which approaches $\frac{1}{2}\Sigma \gamma_i Z_i^2$ + constant as $\sigma_A^2 \to 0$. Hence the critical region corresponding to the one-sided locally best test for $\sigma^2 = 0$ would be $\{Z'\Gamma Z > c\}$. But as $\sigma_A^2 \to 0$, $R(\sigma_A^2) \to \{Z'\Gamma Z > c\}$.

Thus, as usual, the one-sided locally best test is included as a limiting case of the class of β-optimal, level α tests.

At the other extreme one notes that the usual chi-squared test has a critical region of the form $\{Z'Z > c\}$. But as $\sigma_A^2 \to \infty$, $R(\sigma_A^2) \to \{Z'Z > c\}$. Thus the usual chi-squared test is also included as a limiting case of the class of β-optimal, level α tests.

"Y" Critical Regions

Finally we express our critical regions in terms of $Y = <(r_i/n_i-p_i)/(p_iq_i/n_i)^{1/2}>$. We claim that

$$R(\sigma_A^2) = \{\sigma_A^{-2}[Y_1'Q_1Y_1-Y_2'Q_2Y_2] > c\}$$

where

$$Y_1 = Y = <(r_i/n_i-p_i)/(p_iq_i/n_i)^{1/2}>$$

$$Q_1 = I - X(X'X)^{-1}X'$$

$$Y_2 = (\sigma_A^2 ND+I)^{-1/2}Y$$
$$= <(r_i/n_i-p_i)/((\sigma_A^2 n_ip_iq_i+1)p_iq_i/n_i)^{1/2}>$$

$$Q_2 = I - (\sigma_A^2 ND+I)^{-1/2}X(X'(\sigma_A^2 ND+I)^{-1}X)^{-1}X'(\sigma_A^2 ND+I)^{-1/2} .$$

This follows since

$$S(Z, \sigma_A^2) = Z'\Gamma(\sigma_A^2\Gamma+I)^{-1}Z$$

$$= \sigma_A^{-2} Z'[I-(\sigma_A^2\Gamma+I)^{-1}]Z$$

$$= \sigma_A^{-2} Y'C[I-(\sigma_A^2 C'NDC+C'C)^{-1}]C'Y$$

$$= \sigma_A^{-2}[Y'CC'Y-Y'C(C'(\sigma_A^2 ND+I)C)^{-1}C'Y]$$

Now since the range of C is the nullity of X'

$$C(C'(\sigma_A^2 ND+I)C)^{-1}C'$$

$$= (\sigma_A^2 ND+I)^{-1} - (\sigma_A^2 ND+I)^{-1}X(X'(\sigma_A^2 ND+I)^{-1}X'(\sigma_A^2 ND+I)^{-1} .$$

This follows, using Lemma 2.2.6(c) in Rao and Mitra (1971) and basic properties of projections. Thus

$$S(Z, \sigma_A^2)$$

$$= \sigma_A^{-2}[Y'QY-Y'(\sigma_A^2 ND+I)^{-1}Y$$

$$+ Y'(\sigma_A^2 ND+I)^{-1}X(X'(\sigma_A^2 ND+I)^{-1}X)^{-1}X'(\sigma_A^2 ND+I)^{-1}Y]$$

$$= \sigma_A^{-2}[Y'QY-Y_2'Y_2$$

$$+ Y_2'(\sigma_A^2 ND+I)^{-1/2}X(X'(\sigma_A^2 ND+I)^{-1}X)^{-1}X'(\sigma_A^2 ND+I)^{-1/2}Y_2]$$

$$= \sigma_A^{-2}[Y_1'QY_1-Y_2'QY_2]$$

Note further that

$$Y_1'Q_1Y_1 = \min_\beta\{(Y-\mathbf{X}\beta)'(Y-\mathbf{X}\beta)\}$$

and

$$Y_2'Q_2Y_2 = \min_\beta\{(Y-\mathbf{X}\beta)'(\sigma^2 ND+I)^{-1}(Y-\mathbf{X}\beta)\}$$

so that the difference is essentially just the difference in residual sums of squares under the two weighting (covariance) assumptions. This is further justification that the choice of the projection Q (the choice of the "error" space) by which we calculated our residuals was an arbitrary one, and that no information about σ^2 was lost thereby.

The one-sided locally best and usual chi-squared test critical regions can also be expressed in terms of Y as follows.

$$R(0) = \{Y'NDY > c\} \quad \text{(locally best)}$$

$$R(\infty) = \{Y'Y > c\} \quad \text{(chi-squared)}$$

But recall that the $C(\alpha)$ statistic derived during the likelihood analysis was associated with the critical region $\{Y'NDY > c\}$, and the chi-squared test derived with model M_1 was associated with the critical region $\{Y'Y > c\}$. Thus these tests are the same, whether derived in the current "simplified" context or in the original context. But we now have information about the null distribution involved with the one-sided locally best test statistic ($C(\alpha)$ statistic). Since $E[Z'\Gamma Z] = \text{tr}\,\Gamma$ and $V[Z'\Gamma Z] = \text{tr}[2\Gamma^2]$ under our normality assumptions, we can calculate $E[T] = tr[QND-ND]/2$ and $V[T] = tr[(QND)^2]/4$ under normality assumptions, where T is the $C(\alpha)$ test statistic developed in Chapter III.

<u>The Estimation Problem</u>

Let us examine the problem of estimating σ^2 when we have observations Z such that

$$Z \sim N(\phi, \sigma^2\Gamma+I).$$

Thus

$$Z_i^2 \sim (\sigma^2\gamma_i+1)\chi^2(1)$$

so

$$E[Z_i^2] = \sigma^2\gamma_i+1$$

and

$$V[Z_i^2] = 2(\sigma^2\gamma_i+1)^2$$

If we regress Z_i^2-1 on γ_i with weights $W_i = (\sigma^2\gamma_i+1)^2$ we obtain an estimator

$$\hat{\sigma}^2 = \Sigma_{i=1}^n[\gamma_i(Z_i^2-1)/W_i]/\Sigma_i[\gamma_i^2/W_i]$$

which is unbiased and has minimum variance at σ^2 among all unbiased quadratic estimators $(Z'AZ)$ provided that the weights W_i are evaluated at the true σ^2 (see Olsen and Seely (1973)).

We may evaluate the weights at a particular $\sigma^2 = \sigma_A^2$ and then compute the variance of $\hat{\sigma}^2$ to be

$$V_1(\sigma^2;\sigma_A^2) = B\Sigma_{i=1}^n(2\gamma_i^2(\sigma^2\gamma_i+1)^2)/(\sigma_A^2\gamma_i+1)^4 \tag{4.3}$$

where σ^2 is the "true" value and

$$B = [1/\Sigma_{i=1}^n(\gamma_i^2/(\sigma_A^2\gamma_i+1)^2]^2$$

Note that when $\sigma_A^2 = \sigma^2 =$ "true" σ^2 this simplifies to

$$V(\sigma^2;\sigma^2) = 2/\Sigma_{i=1}^n[\gamma_i^2/(\sigma^2\gamma_i+1)^2] .$$

Further note that the maximum likelihood estimator $\hat{\rho}$ of σ^2 satisfies the equation

$$\Sigma_{i=1}^n \frac{\hat{\rho}\gamma_i^2}{(\hat{\rho}\gamma_i+1)^2} = \Sigma_{i=1}^n \frac{\gamma_i(Z_i^2-1)}{(\hat{\rho}\gamma_i+1)^2}$$

Thus

$$\hat{\rho} = \Sigma_i \frac{\gamma_i(Z_i^2-1)}{(\hat{\rho}\gamma_i+1)^2} \Big/ \Sigma_i \frac{\gamma_i^2}{(\hat{\rho}\gamma_i+1)^2}$$

so that, in particular, if we estimate $\hat{\sigma}^2$ as above, with weights evaluated at some σ_A^2 and then use this $\hat{\sigma}^2$ in the weights to get a new $\hat{\sigma}^2$, use this new $\hat{\sigma}^2$ in the weights to get yet another $\hat{\sigma}^2$, etc., we eventually would converge to a solution of the likelihood equation. Finally, note that the maximum likelihood estimate has asymptotic variance

$$[E(-\partial^2 \ln f/(\partial\rho)^2)]^{-1} \simeq [2/\Sigma_i\ \gamma_i^2(\hat{\rho}\gamma_i+1)^2]$$

which corresponds to $V(\hat{\rho};\hat{\rho})$.

The Chi-Squared Estimator

An alternative way to estimate σ^2 is to find an unbiased estimator based on the usual chi-squared test statistic $Z'Z$, for which $E(Z'Z) = \text{tr Cov}[Z] = \sigma^2\Sigma_{i=1}^n\ \gamma_i + (n-(m+1))$. Thus we may write

$$\hat{\sigma}^2 = [Z'Z-(n-(m+1))]/\Sigma\gamma_i$$

with variance, as a function of the true σ^2,

$$V_2(\sigma^2) = 2B\Sigma_{i=1}^n\ (1+\sigma^2\gamma_i)^2 \qquad (4.4)$$

where

$$B = [1/\Sigma\gamma_i]^2$$

Expressing this estimator in terms of Y, $\hat{\sigma}^2 = [Y'Y-(n-(m+1))]/\text{tr}[QND]$, we see that it is a function of Finney's heterogeneity factor. Thus it is, in a sense, the most natural way to implement model M_2 when model M_1 fails to be adequate. Unfortunately, under mild conditions on $<\gamma_i>$, there is always a "regression" type estimator with variance uniformly (over σ^2) less than the variance of this "chi-squared" type estimator. That is, there exists a σ^2_A such that $V_1(\sigma^2; \sigma^2_A) < V_2(\sigma^2)$ for all $\sigma^2 > 0$.

To see this, we define

$$U = <Z_i^2 - 1>$$
$$H = <\gamma_i>$$
$$H* = <\gamma_i^2>$$
$$\Gamma = \text{diag} \ll\gamma_i\gg$$

then our regression situation can be expressed

$$U = H\sigma^2 + e$$

where

$$e \sim WS(\phi, 2\sigma^4\Gamma^2 + 4\sigma^2\Gamma + 2I)$$

Further define $[\mathcal{V}]$ to be the smallest closed convex cone in the space of symmetric matrices containing $V(\sigma^2) = 2\sigma^4\Gamma^2 + 4\sigma^2\Gamma + 2I$ for all $\sigma^2 \geq 0$. By Olsen, Seely and Birkes (1975) Corollary 3.12, an unbiased linear function σ^2 of $\{Z_i\}$ is admissible if and only if it is a solution of

$$H'V^{-1}H\sigma^2 = H'V^{-1}U$$

for some V in $[\mathcal{U}]$. However our chi-squared estimator can be expressed

$$\hat{\sigma}^2 = [\Sigma_{i=1}^n (Z_i^2 - 1)]/\Sigma_{i=1}^n \gamma_i$$

$$= H'\Gamma^{-1}U/H'\Gamma^{-1}H$$

Now suppose that $V \in [\mathcal{U}]$ and note that $[\mathcal{U}] \subset \{a_1\Gamma^2 + a_2\Gamma + a_3 I : a_i \geq 0, i = 1, 2, 3\}$ so that $V = a_1\Gamma^2 + a_2\Gamma + a_3 I$ for at least one triple as above. If, for all U,

$$H'V^{-1}U/H'V^{-1}H = H'\Gamma^{-1}U/H'\Gamma^{-1}H$$

then

$$H'V^{-1} = kH'\Gamma^{-1}$$

where

$$k = H'V^{-1}H/H'\Gamma^{-1}H .$$

It follows that

$$\begin{aligned} H' &= kH'\Gamma^{-1}V \\ &= kH'\Gamma^{-1}[a_1\Gamma^2 + a_2\Gamma + a_3 I] \\ &= kH'[a_1\Gamma + a_2 I + a_3\Gamma^{-1}] \\ &= k[a_1 H'\Gamma + a_2 H' + a_3 H'\Gamma^{-1}] \end{aligned}$$

and hence

$$H = k[a_1H^* + a_2H + a_3 1]$$

or

$$a_1H^* + (a_2 - 1)H + a_3 1 = 0$$

Now we make the assumptions that $n \geq 3$ and $<\gamma_i>$ is such that

$$\det[H^* \vdots H \vdots 1] \neq 0.$$

This implies that the only solution to the above is $a_1 = 0$, $a_2 = 1$, $a_3 = 0$ so that $V = \Gamma$. But it is clear that $\Gamma \notin W = \{2\sigma^4\Gamma^2 + 4\sigma^2\Gamma + 2I : \sigma^2 \geq 0\}$ and by arguments similar to those used by Olsen, Seely and Birkes (1975) concerning Theorem 4.3, Γ is not in the closed convex cone $[\bigcup]$ generated by W. Thus the chi-squared estimator is not admissible. Its usefulness however remains to be judged.

"Y" Estimators

The estimator

$$\hat{\sigma}^2 = \Sigma[\gamma_i(Z_i^2 - 1)/W_i]/\Sigma[\gamma_i/W_i]$$

where $W_i = (\sigma_A^2\gamma_i - 1)^2$ can be expressed

$$\hat{\sigma}^2 = \frac{Z'\Gamma(\sigma_A^2\Gamma+I)^{-2}Z - 1'\Gamma(\sigma_A^2\Gamma+I)^{-2}1}{1'\Gamma^2(\sigma_A^2\Gamma+I)^{-2}1}$$

$$= \frac{P_1 - P_2}{P_3}$$

Now using the reation

$$\frac{2\sigma^2\gamma}{(1+\sigma^2\gamma)^2} = 1 - \frac{1}{(1+\sigma^2\gamma)^2} - \frac{\sigma^4\gamma^2}{(1+\sigma^2\gamma)^2}$$

we have

$$2\sigma^2 Z'\Gamma(I+\sigma^2\Gamma)^{-2}Z$$

$$= Z'Z - Z'(I+\sigma^2\Gamma)^{-2}Z - \sigma^4 Z'\Gamma(1+\sigma^2\Gamma)^{-2}\Gamma Z$$

But

$$Z'\Gamma(I+\sigma^2\Gamma)^{-2}\Gamma Z$$

$$= Y'NDC(C'C+\sigma^2C'NDC)^{-2}C'NDY$$

$$= Y'NDC(C'(I+2\sigma^2ND+\sigma^4NDQND)C)^{-1}C'NDY$$

By Lemma 2.2.6(c) in Rao and Mitra (1971) and basic properties of projections we see that $C(C'MC)^{-1}C = M^{-1} - M^{-1}X(X'M^{-1}X)^{-1}X'M^{-1}$ so if $M = I + 2\sigma^2ND + \sigma^4NDQND$ then

$$Z'\Gamma(I+\sigma^2\Gamma)^{-2}\Gamma Z = Y'ND[M^{-1} - M^{-1}X(X'M^{-1}X)^{-1}X'M^{-1}]NDY$$

Similarly

$$Z'(I+\sigma^2\Gamma)^{-2}Z = Y'[M^{-1}-M^{-1}X(X'M^{-1}X)^{-1}X'M^{-1}]Y$$

and

$$Z'Z = Y'CC'Y = Y'Y$$

Thus

$$P_1 = \frac{1}{2\sigma^2}[Y'Y-Y'[M^{-1}-M^{-1}X(X'M^{-1}X)^{-1}X'M^{-1}]Y$$
$$-\sigma^4 Y'ND[M^{-1}-M^{-1}X(X'M^{-1}X)^{-1}X'M^{-1}]NDY]$$

P_2 and P_3 can be similarly reduced to terms involving Y, N, D, Q, X and σ^2.

V. ASYMPTOTIC PROPERTIES

In this chapter we discuss the asymptotic distribution of the maximum likelihood estimator θ for our context. This context is non-standard in two respects. First we must contend with observations which are independent but not identically distributed (i.n.i.d.). Secondly, we must cope with a parameter which may lie on the boundary of the parameter space.

The first of these conditions has been considered by Hoadley (1971), and Bradley and Gart (1962) among others. The second, by Chant (1974) and Moran (1971b) among others. However we know of no place in the literature where these are addressed simultaneously for our context.

We will first establish further notation and definitions, then display special properties of our likelihood function and assumptions about the sequence of design matrices involved, and finally discuss the distributional results which would be used in a purely likelihood analysis of data under model M_2.

Notation

We make the following definitions:

ϕ_n = log likelihood associated with n observations

$\hat{\theta}_n$ = m.l.e. associated with n observations

$$z_n = n^{1/2}(\hat{\theta}_n - \theta)$$

$$\bar{g}_n(\theta; R) = g_n(\theta; R)/n$$

$$\bar{G}_n(\theta) = E_\theta[-G_n(\theta; R)]/n$$

Note that since

$$E_\theta[-G_n(\theta; R)] = E_\theta[g_n(\theta; R)g_n'(\theta; R)]$$

we have

$$\bar{G}_n(\theta) = nE_\theta[\bar{g}_n(\theta; R)\bar{g}_n'(\theta; R)]$$

$$= \mathrm{Cov}_\theta[n^{1/2}\bar{g}_n(\theta; R)]$$

Further define

$$f^{(r,s,t)} = \partial^3 f/\partial\theta_r \partial\theta_s \partial\theta_t$$

$$= \int h[(r-np)^3 - npq(3[r-np]+q-p)][\partial\lambda/\partial\theta_r]$$

$$\times [\partial\lambda/\partial\theta_s][\partial\lambda/\partial\theta_t]dF(e)$$

so that

$$\partial^3\phi/\partial\theta_r\partial\theta_s\partial\theta_t = \Sigma_{i=1}^n \{f_i^2 f_i^{(r,s,t)} - f_i[f_i^{(r)}f_i^{(s,t)} + f_i^{(s)}f_i^{(r,t)} + f_i^{(t)}f_i^{(r,s)}]$$

$$+ 2f_i^{(r)}f_i^{(s)}f_i^{(t)}\}/f_i^3$$

We will say that a function $f(i, r, \theta)$ is "uniformly continuous in θ over θ, r, and i" if $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$|\theta_1-\theta_2| < \delta$ implies $|f(i,r,\theta_1)-f(i,r,\theta_2)| < \epsilon$ for all i and r, where δ is independent of θ_1, θ_2, r and i.

Conditions

We will make the following assumptions:

A1. The group sizes are bounded, say $1 \leq n_i \leq K_1$.

A2. The parameter space Ω is a closed, bounded subset of R^{m+2}, say $0 \leq \theta_1 \leq K_2$ and $-K_2 \leq \theta_s \leq K_2$, for $s = 2, \ldots, m+2$.

A3. We have a sequence of "design" matrices $\{A_{(i)}\}$, where $A_{(i+1)}$ differs from $A_{(i)}$ in having one more row appended, and the entries of all these matrices are bounded, say $-K_2 \leq a_i^j \leq K_2$

A4. This sequence $\{A_{(i)}\}$ is such that there is an N_0 such that $\forall \rho > 0, \exists \eta > 0$ which is independent of θ_1 and θ_2 such that $|\theta_1-\theta_2| \geq \rho$ and $n \geq N_0$ implies

$$\frac{1}{n} \Sigma_{i=1}^{n} E_{\theta_1} [\log f_i(r_i;\theta_1) - \log f_i(r_i;\theta_2)] > \eta .$$

A5. This sequence $\{A_{(i)}\}$ is such that for each $\theta \in \Omega$ $\bar{G}_n(\theta) \to \bar{G}(\theta)$, where $\bar{G}(\theta)$ is positive definite with finite determinant.

Note 1: Under these conditions f_i is bounded, both above and away from zero, say $0 < K_3 \leq f_i \leq \frac{1}{4}$. To see this we note that conditions A2 and A3 imply that $|A_i\beta| \leq K_4$ where K_4 depends on K_2 and m but not on i or n. Thus

$$|E(\lambda_i)| = E(A_i\beta+\sigma e_i)| = |A_i\beta| \leq K_4$$

Now

$$h_i = p_i^{r_i} q_i^{n_i - r_i} \leq p_i q_i \leq \frac{1}{4}$$

so

$$f_i = E[h_i] \leq \frac{1}{4}$$

Further since $0 \leq r_i \leq n_i \leq K_1$ we have that $h_i \geq [p_i q_i]^{K_1}$. But

$$p_i = 1/(1+\exp(-\lambda_i))$$

$$\geq 1/(1+\exp(K_4 - \sigma e))$$

and

$$q_i \geq 1/(1+\exp(K_4 + \sigma e))$$

so

$$h_i \geq \begin{cases} p_i^{2K_1} & \text{if} \quad p_i \leq q_i \\ q_i^{2K_1} & \text{if} \quad p_i > q_i \end{cases}$$

However, when the distribution of e is symmetric, the expectation is the same in either case. Thus

$$f_i = E[h_i] \geq E[(1+\exp(K_4+\sigma e))^{-2K_1}]$$

$$= E[(1+\exp(K_4-\sigma e))^{-2K_1}]$$

Now

$$\partial E[1+\exp(K_4+\sigma e))^{-2K_1}]/\partial\sigma$$

$$= -2K_1 \int_{-\infty}^{\infty} (1+\exp(K_4+\sigma e)^{-2K_1-1} \exp(K_4+\sigma e) e dF(e)$$

$$= -2K_1 \int_0^{\infty} [(1+\exp(K_4+\sigma e))^{-2K_1-1} \exp(K_4+\sigma e)$$

$$- (1+\exp(K_4-\sigma e))^{-2K_1-1} \exp(K_4-\sigma e)] e dF(e)$$

Define $H(x) = (1+\exp(K_4+x))^{-2K_1-1} \exp(K_4+X)$. We see that

$$\frac{dH}{dx} = (1+\exp(K_4+x))^{-2K_1} [\frac{\exp(K_4+x)}{1+\exp(K_4+x)}][1-(2K_1+1)\frac{\exp(K_4+x)}{1+\exp(K_4+x)}]$$

Since the first two terms are positive but the last is negative dH/dx is always negative, and hence H is a monotone decreasing function. In particular $H(\sigma e) - H(-\sigma e) < 0$ when $\sigma e > 0$.

Therefore $\partial E[(1+\exp(K_4+\sigma e))^{-2K_1}]/\partial\sigma$ is positive for all $\sigma > 0$, and hence monotone increasing for $\sigma > 0$. Since $E[(1+\exp(K_4+\sigma e))^{-2K_1}]$ is a continuous function at σ over $\sigma \geq 0$, it takes its minimum value at $\sigma = 0$. Thus

$$f_i \geq (1+\exp(K_4))^{-2K_1} = K_3 > 0$$

for all $\sigma \geq 0$.

Distribution of m.l.e.

The asymptotic distribution of our maximum likelihood estimator $\hat{\theta}$ will be established in several steps. The first of these steps is a generalization of Moran's (1971a) proof of the uniform consistency of $\hat{\theta}$. Our definitions and A2 verify Moran's assumptions 1-3. A2 and Note 1 enable us to show that ϕ_n is uniformly bounded and uniformly continuous in θ over Θ and R. This and A4 enable us to prove his Theorem 1 for our context. This theorem states that for $\rho > 0$, and Ω_2 any closed set in Ω whose distance from θ_1 is greater than ρ, we have given any $\epsilon_1 > 0$, $\epsilon_2 > 0$ there exists N_0 independent of θ_1 so that for all $n > N_0$ we have

$$\text{Prob}_{\theta_1}\left[\frac{\sup f_1(r_1,\theta)\cdots f_n(r_n,\theta)}{f_1(r_1,\theta_1)\cdots f_n(r_n,\theta_1)} > \epsilon_1\right] < \epsilon_2$$

where the supremum is taken over all values of θ in Ω_2. Having established this theorem, uniform consistency follows directly.

Next we must show that the gradient, $n^{1/2}\bar{g}_n$ (evaluated at the "true" value θ of the parameter), is distributed multivariate normal $N(\phi, \bar{G}(\theta))$. We use the uniform bound on $\bar{g}_n$ and A5 to apply Liapounov's multivariate central limit theorem, as proved in Hoadley (1971).

Finally we must infer the distribution of z_n from the distribution of the $n^{1/2}\bar{g}_n$. This is considered in two cases, where θ is not on the boundary (we can assume that $\hat{\theta}$ is in an open neighborhood of θ), and where θ is on the boundary of Ω (that is the "true" value of σ^2 is zero). Both cases depend on the asymptotic validity of the usual Taylor's approximation.

$$\bar{g}(\hat{\theta}; R) \doteq \bar{g}(\theta; R) - (\hat{\theta}-\theta)'\bar{G}(\theta) . \tag{5.1}$$

One of the early expositions of this relationship and its validity was in Wald (1943). Under a somewhat different set of assumptions Chanda (1954) also established this result. Moran (1971b) extended Wald's work to include the situation where the true parameter point lies on the boundary of Ω. Bradey and Gart (1962) extended Chanda's work to associated populations (the independent not identically distributed case).

In the first case $\bar{g}(\hat{\theta}; R) = \phi$, so from (5.1)

$$(\hat{\theta}-\theta) \doteq [\bar{G}(\theta)]^{-1}\bar{g}(\theta; R)$$

and $n^{1/2}\bar{g}(\theta; R)$ has been established to be asymptotically normal. This along with standard results (Chiang (1956) pg 338) can be used to show that asymptotically $z_n \sim N(\phi, [\bar{G}(\theta)]^{-1})$.

In the second case, where the true value of θ is on the boundary, we extend Chant's work (1974). At the maximum, $\bar{g}_n(\hat{\theta}) = (b, 0, \ldots, 0)'$ where b is non-positive if $\hat{\theta}_1 = 0$ and zero if $\hat{\theta}_1 > 0$. Conditioning on these two situations the following result is obtained

$$\Pr\{z_n \leq u; \theta\} \to \frac{1}{2} F_1(u; \theta) + \frac{1}{2} F_2(u; \theta)$$

where:

i) F_1 is a (m+2)-dimensional multivariate distribution defined on the region $u_1 > 0$, $-\infty < u_s < \infty$ $(s = 2, \ldots, m+2)$, and having in this region a probability density equal to twice the density of a multivariate normal distribution with means zero, and covariance $[\bar{G}(\theta)]^{-1}$.

ii) F_2 is a (m+1)-dimensional distribution concentrated on $u_1 = 0$, $-\infty < u_s < \infty$ $(s = 2, \ldots, m+2)$ and having in this region a multivariate normal distribution with means zero and covariance $[\bar{G}(\theta)_1]^{-1}$, where $\bar{G}(\theta)_1$ is that minor of $\bar{G}$ obtained by striking out the first row and first column.

This last result is of great value when evaluating the conclusions reached during a likelihood analysis in the case that M_1 has been improperly rejected. In this case the asymptotic marginal distribution of $\hat{\beta}$ would not be normal.

VI. SUMMARY, AND AN EXAMPLE

In previous chapters we have developed a model more general than the traditional one, and considered theoretical aspects of a variety of tests and estimators associated with it. In this chapter we would like to discuss the practical differences one could expect to find i) between the powers of the various tests, ii) between the variances of the various estimators, and iii) between the covariance matrices associated by the two models with estimates of β. We will do this in the context of example 8, pages 72-76 of Finney (1971).

The Example

The results were originally obtained by Bosvine and discussed by Bliss (1940). They concern the toxicity of ethylene oxide to the grain beetle, Calandra granaria. The data are shown in Table 1, x being the log base 10 of the concentration of ethylene oxide in mg/100 ml. Logit analysis led to the relation $\lambda = -3.443 + 14.440x$, where λ is the logit of p. The chi-squared goodness of fit statistic was 33.2445 on eight degrees of freedom. Thus Finney's heterogeneity factor was 33.2445/8 = 4.16. Maximum likelihood estimation, under M_2, led to $\lambda = -4.257 + 17.633x$ with $\hat{\sigma}^2 = .707$. The chi-squared estimate of σ^2 was .663.

Table 1. Results of ethylene oxide-*Calandra granaria* experiment.

x	n	r
0.394	30	23
0.391	30	30
0.362	31	29
0.322	30	22
0.314	26	23
0.260	27	7
0.225	31	12
0.199	30	17
0.167	31	10
0.033	24	0

Power and Variance Calculations

Table 2 displays how several "locally best" tests of $\sigma^2 = 0$ vs $\sigma^2 > 0$ compare for this example. The comparison is in terms of the power of each of these tests against alternatives ranging from $\sigma^2 = 0$ to $\sigma^2 = 1.5$. A Fortran program called TABLE, based on (4.2) was used for these calculations, and is displayed in the Appendix.

Table 3 displays how several "locally best" estimators and the chi-squared estimator compare for this example. The comparison is in terms of the variance of these estimators at "true" σ^2 values ranging from 0 to 1.5. The variance of each of these estimators can be expressed as a quadratic function of the "true" value of σ^2, as in (4.3) and (4.4). The coefficients of these particular estimators are displayed in Table 4. Note that the chi-squared estimator is

Table 2. Powers of "locally best" and chi-squared tests against various alternatives.

"Alt." σ^2	$\sigma_A^2 = 0$	$\sigma_A^2 = .7$	$\sigma_A^2 = 1.4$	$\sigma_A^2 = \infty$
0.0	.050	.050	.050	.050
0.1	.242	.239	.237	.232
0.2	.443	.446	.443	.436
0.3	.598	.606	.603	.597
0.4	.707	.718	.716	.712
0.5	.783	.795	.794	.791
0.6	.837	.849	.848	.846
0.7	.875	.886	.886	.884
0.8	.903	.913	.913	.912
0.9	.923	.933	.932	.932
1.0	.939	.947	.947	.946
1.1	.951	.958	.958	.957
1.2	.960	.966	.966	.966
1.3	.967	.972	.972	.972
1.4	972	.977	.977	.977
1.5	.977	.981	.981	.981

Table 3. Variances of chi-squared and locally best estimators.

"True" σ^2	Chi-Squared	$\sigma_A^2 = 0$	$\sigma_A^2 = .24$	$\sigma_A^2 = .7$	$\sigma_A^2 = \infty$
0.0	.0110	.0096	.0110	.0136	.0200
0.1	.0244	.0247	.0244	.0271	.0352
0.2	.0436	.0471	.0435	.0458	.0554
0.3	.0685	.0767	.0684	.0699	.0807
0.4	.0991	.1136	.0990	.0991	.1109
0.5	.1355	.1577	.1353	.1336	.1462
0.6	.1777	.2092	.1774	.1733	.1864
0.7	.2256	.2678	.2252	.2183	.2316
0.8	.2792	.3338	.2787	.2685	.2819
0.9	.3386	.4070	.3380	.3239	.3371
1.0	.4038	.4874	.4030	.3846	.3973
1.2	.5514	.6701	.5502	.5218	.5328
1.3	.6338	.7723	.6324	.5982	.6080
1.4	.7220	.8818	.7203	.6798	.6883
1.5	.8159	.9985	.8140	.7667	.7735

dominated by the estimator best at .24. In fact, it is dominated by estimators best at values from .233 to .245.

Table 4. Coefficients of variance functions.

Coel.	Chi-Square	0	24	.7	∞
a_1	.0110	.0096	.0110	.0156	.0200
a_2	.1051	.1149	.1052	.1090	.1274
a_3	.2877	.3629	.2867	.2621	.2500

For this example the roots of QNDQ used in the above calculations were {7.432, 6.664, 6.254, 5.127, 4.563, 3.408, 2.641, 1.970}.

Examining these tables, and similar ones for other data sets, one comes to feel that although the chi-squared test is not best against alternatives most likely to be of interest and the chi-squared estimator is not even admissible, they are not "significantly" less powerful or less efficient. Thus, in view of the ease with which they can be implemented, we must recommend their use for "reasonable" data sets, that is data where few if any of the $\lambda_i = A_i\hat{\beta}^0$ are less than -3 or greater than 3. If the data is not reasonable in this sense then we must recommend either the test and estimator best at $\sigma^2 = 0$ or judicious "trimming" as discussed by Finney (1971) to make the chi-squared test and estimator more stable.

$\underline{Cov(\beta)\ for\ M_1\ and\ M_2}$

If model M_1 is used to estimate β^0, and there is no heterogeneity, then we have

$$Cov_1(\hat{\beta}^0) = (X'X)^{-1} .$$

If model M_1 is used to estimate β^0, but there is heterogeneity so that M_2 is appropriate, then we have the true covariance

$$Cov_2(\hat{\beta}^0) = (X'X)^{-1}X'[I+\sigma^2 ND]X(X'X)^{-1}$$

Finney (1971) suggests in this case approximating Cov_2 by a "corrected Cov_1," namely $h\ Cov_1$, where h is the heterogeneity factor.

If the heterogeneity is detected, σ^2 estimated and a least squares estimator $\hat{\beta}^+$ calculated with M_2 weights, then we have

$$Cov_3(\hat{\beta}^+) = (X'(I+\hat{\sigma}^2 D)^{-1}X)^{-1}$$

Finally if heterogeneity is detected and an M_2 maximum likelihood estimator $\hat{\beta}$ calculated, then we have

$$Cov_4(\hat{\beta}) = [G(\theta;R)]_1^{-1}$$

Now we can compare Finney's "corrected Cov_1," where for h is substituted its expected value $1 + \sigma^2 \bar{d}$, with the true variance Cov_2,

$$Cov_2 - h\, Cov_1 = \sigma^2[(X'X)^{-1}X'(ND - \bar{d}I)X(X'X)^{-1}]$$

where $\bar{d} = tr[QNDQ]/(n-(m+2))$. This shows us that when the entries of ND are all roughly the same then Finney's heterogeneity factor h will work well, but if the entries of ND are very diverse then it will be necessary to go to a more general model to accurately calculate $Cov(\hat{\beta})$.

For our example we have $\bar{d} = 4.75737$,

$$(X'X)^{-1}E(h) = \begin{bmatrix} 1.068 & -3.700 \\ -3.700 & 14.028 \end{bmatrix}$$

$$(X'X)^{-1}X(I+\hat{\sigma}^2D)X(X'X)^{-1} = \begin{bmatrix} 1.032 & -3.359 \\ -3.359 & 12.145 \end{bmatrix}$$

$$(X'(I+\hat{\sigma}^2D)^{-1}X)^{-1} = \begin{bmatrix} 0.853 & -2.791 \\ -2.791 & 10.293 \end{bmatrix}$$

and

$$[G(\hat{\theta};R)]_1^{-1} = \begin{bmatrix} 1.204 & -4.060 \\ -4.060 & 14.981 \end{bmatrix}$$

Recall that, rather than actually using Finney's h, we have used the expected value of h (based on maximum likelihood estimate $\hat{\sigma}^2 = .7073$) so this result will be more reasonably compared with the others.

For this data, Finney's h seems to be quite reasonable. Notice however that if we were to estimate σ^2, and perform another least squares iteration with corrected weights, the precision of the estimator seems to increase substantially. The last calculation might imply that Cov_4 is more conservative, or that estimation by likelihood procedures is less precise.

Certainly no hard and fast rules can be formulated by looking at a few examples. However, it seems that the chi-squared test (perhaps trimmed), the chi-squared estimator and weighted least squares in the context of model M_2 hold great promise to be the most dependable and easily implemented of the possibilities we have examined.

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I.A. (1966) Handbook of Mathematical Functions. National Bureau of Standards, Applied Mathematics Series-55.

Anderson, T.W. (1971) The Statistical Analysis of Time Series (pg. 611). John Wiley and Sons, New York.

Berkson, J. (1951) Why I Prefer Logits to Probits. Biometrics 7, 327-339.

Bliss, C.I. (1940) The relation between exposure time, concentration and toxicity in experiments on insecticides Ann. Ent. Soc. Amer. 33, 721-66.

Bradley, E.L. (1973) The equivalence of maximum likelihood and weighted least squares estimates in the exponential family. Journal of the American Statistical Association 68, 199-200.

Bradley, R.A. and Gart, J.J. (1962) The asymptotic properties of ML estimators when sampling from associated populations. Biometrika 49, 205-214.

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

Chant, D. (1974) On asymptotic tests of composite hypotheses in nonstandard conditions. Biometrika 61, 291-298.

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

Clutton-Brock, M. (1967) Likelihood distributions for estimating functions when both variables are subject to error. Technometrics 9, 261-269.

Cornfield, J. (1952) Measurement and comparison of toxicities: the quantal response. Chapter 25 of Statistics and Mathematics in Biology, Hafner, New York (pgs. 327-344).

Cox, D.R. (1970) The analysis of binary data. Methuen, London.

David, F.N. (1951) Probability theory for statistical methods. Cambridge University Press.

Davidon, W.C. (1959) Variable metric method for minimization. Argonne National Laboratory ANL 5990 Rev.

Davies, R.B. (1969) Beta-optimal tests and an application to the summary evaluation of experiments. Journal of the Royal Statistical Society-Series B 31, 524-538.

Dews, P.B. and Berkson, J. (1952) On the error of bioassay with quantal response. Chapter 27 of Statistics and Mathematics in Biology, Hafner, New York (pgs. 361-370).

Finney, D.J. (1964) Statistical methods in biological assay. Griffin, London.

Finney, D.J. (1971) Probit Analysis (3rd ed.). Cambridge University Press.

Hoadley, B. (1971) Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. Annals of Mathematical Statistics 42, 1977-1991.

Marquardt, D.W. (1970) Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. Technometrics 12, 591-612.

Moran, P.A.P. (1970) On asymptotically optimal tests of composite hypotheses. Biometrika 57, 47-55.

Moran, P.A.P. (1971a) The uniform consistency of maximum likelihood estimators. Proceedings, Cambridge Philosophical Society 70, 435-439.

Moran, P.A.P. (1971b) Maximum likelihood estimation in non-standard conditions. Proceedings, Cambridge Philosophical Society 70, 441-450.

Murray, W. (1972) Numerical methods for unconstrained optimization. Academic Press, New York.

Olsen, A.R. and Seely, J. (1973) Invariant quadratic unbiased estimation for two variance components. Technical Report #37. Department of Statistics, Oregon State University.

Olsen, A R., Seely, J., and Birkes, D. (1975) Invariant quadratic unbiased estimation for two variance components. To appear in Annals of Mathematical Statistics.

Patwary, K.M. and Haley, K D.C. (1967) Analysis of quantal response assays with dosage errors. Biometrics 23, 747-760.

Rao, C.R. and Mitra, S.K. (1971) Generalized inverse of matrices and its applications. John Wiley and Sons, New York.

Seely, J. (1972) Completeness for a family of multivariate normal distributions. Annals of Mathematical Statistics 43, 1644-1647.

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

APPENDIX

In this appendix are displayed two of the programs used in this work. The first uses the modified Newton-Raphson algorithm discussed in the body of this thesis to maximize the M_2 likelihood, where $\sigma^2 > 0$. The second was used to calculate the powers which compared various β-optimal, level α tests.

```
      DEFINE COMMON
      COMMON U(20),W(20),X(6,30),B1(7),B2(7),Y(30),N(30)
     1,G(7),H(28),D(7),NR,NC,NC1,L1,L2,L3
      REAL L1, L2, L3, N
      END
      PROGRAM HETLIK
      INCLUDE COMMON
      DIMENSION IFOR(16)
      READ (2,1000) (IFOR(I),I=1,16)
      READ(2,1010) NR,NC
      DO 10 I=1,NR
   10 READ(1,IFOR) (Y(I),N(I),(X(J,I),J=1,NC))
C
      READ(3,1020) (U(I),W(I),I=1,10)
      DO 11 I=1,10
      W(I)=W(I)*.56418955835
      U(I)=U(I)*1.414213562
      II=I+10
      U(II)=-U(I)
   11 W(II)=W(I)
C
      NC1=NC+1
      READ(2,1030) (B1(J),J=1,NC1)
      L3=-10E10
      EPS=10E-6
      CALL LGH
C
   12 WRITE(61,1040) L1
      WRITE(61,1050)(B1(J),G(J),J=1,NC1)
C
      CALL SINV(H,NC1,EPS,IER,DET)
      IF(IER) 13,14,13
C
   13 WRITE(61,1070)
      CALL EXIT
C
   14 IF((L1-L3).LT..02) GO TO 72
      CALL MPRD(H,G,D,NC1,NC1,1,0,1)
C
      STEP = 0.0
      KEND = 0
      L2 = L1
      L3 = L1
      G2 = 0.0
      DO 20 K=1,NC1
      B2(K) = B1(K)
   20 G2 = G2 + G(K)*D(K)
      IF(G2) 71,71,22
C
   22 AL = 10.0/G2
      IF(AL.LT..75) AL = .75
      IF(AL.GT.1.5) AL = 1.5
C
```

```
   24 AA = 1.0
      ALTEST = -.99*B2(NC1)/D(NC1)
      IF(AL.GT.ALTEST.AND.D(NC1).LT.0) AL = ALTEST
      DO 26 J=1,NC1
   26 B1(J) = B2(J) + AL*D(J)
   28 CALL LGH
      WRITE (61,1060) L1
      G1 = 0.
      DO 30 K=1,NC1
   30 G1 = G1 + G(K)*D(K)
      IF(G1) 36,40,32
C
   32 IF(L1-L2.LE.0.0) GO TO 44
      KEND = KEND + 1
      IF(KEND.GT.5) GO TO 42
      L2 = L1
      G2 = G1
      ALTEST = -.99*B1(NC1)/D(NC1)
      IF(AL.GT.ALTEST.AND.D(NC1).LT.0) AL = ALTEST
      DO 34 K=1,NC1
      B2(K) = B1(K)
   34 B1(K) = B2(K) + AL*D(K)
      STEP = STEP + AL
      GO TO 28
C
   36 ZA = 3.*(L2 - L1)/AL + G2 + G1
      QA = -SQRT(ZA*ZA - G2*G1)
      AA = 1. - ((G1 + QA - ZA)/(G1 - G2 + 2.*QA))
      DO 38 K=1,NC1
   38 B1(K) = B2(K) + AL*AA*D(K)
      CALL LGH
C
   40 IF(L1-L2.LE.0.0) GO TO 44
   42 STEP = STEP + AL*AA
      WRITE(61,1080) STEP
      GO TO 12
C
   44 AL = 0.2*AL
      GO TO 24
C
   71 WRITE(61,1100)
   72 WRITE(61,1110)
      JJ=1
      DO 73 J=1,NC1
      DO 73 K=1,J
      WRITE(61,1090) J,K,H(JJ)
   73 JJ=JJ+1
      CALL EXIT
C
```

```
 1000 FORMAT(16A4)
 1010 FORMAT(2I3)
 1020 FORMAT(2E18.10)
 1030 FORMAT(7F10.)
 1040 FORMAT(/,2X,≠L = ≠,F12.3)
 1050 FORMAT(/,(1X,F12.6,2X,E10.3))
 1060 FORMAT(/,2X,≠(L= ≠,F12.3,≠)≠)
 1070 FORMAT(/,1X,≠PROBLEM IN INVERSE≠)
 1080 FORMAT(/,2X,≠STEP LENGTH =≠,F10.3)
 1090 FORMAT(1X,2I2,2X,F12.6)
 1100 FORMAT(/,2X,≠DIRECTIONAL DERIVATIVE .LE. ZERO≠)
 1110 FORMAT(2X,/)
      END
C
C
      SUBROUTINE LGH
      INCLUDE COMMON
      REAL NT
      L1=0.
      DO 10 I=1,NC1
   10 G(I)=0.
      NH = ((NC + 1)*(NC + 2)/2)
      DO 12 I=1,NH
   12 H(I)=0.
      B1(NC1) = SQRT(B1(NC1))
C
      DO 22 I=1,NR
      Z=0.0
      DO 14 J=1,NC
   14 Z=Z+X(J,I)*B1(J)
      YT=Y(I)
      NT=N(I)
      S=S1=S1U=S2=S2U=S2U2=0.
C
      DO 16 L=1,20
      UT=U(L)
      ZT=Z+B1(NC1)*UT
      CALL LKLIHD(ZT,YT,NT,P,Q,F)
      DY = YT - NT*P
      D2Y = DY*DY - NT*P*Q
      F = F*W(L)
      S=S+F
      S1=S1+F*DY
      S1U=S1U+F*DY*UT
      S2=S2+F*D2Y
      S2U=S2U+F*D2Y*UT
      S2U2=S2U2+F*D2Y*UT*UT
   16 CONTINUE
C
```

```
      L1=L1+ALOG(S)
      F1=S1/S
      F2=S2/S-F1*F1
      JJ=1
      DO 18 J=1,NC
      XJ=X(J,I)
      G(J)=G(J)+XJ*F1
      DO 18 K=1,J
      H(JJ)=H(JJ)-XJ*X(K,I)*F2
   18 JJ=JJ+1
      G(NC1)=G(NC1)+S1U/S
      F3=S2U/S-S1*S1U/S**2
      DO 20 J=1,NC
      H(JJ)=H(JJ)-X(J,I)*F3
   20 JJ=JJ+1
   22 H(JJ)=H(JJ)-S2U2/S+(S1U/S)**2
C
      G(NC1) = G(NC1)*.5/B1(NC1)
      NS = NC*NC1/2
      DO 26 J=1,NC
      NS = NS + 1
   26 H(NS) = H(NS)*.5/B1(NC1)
      NS = NS + 1
      H(NS) = (H(NS) - 2.*G(NC1)) * .25/(B1(NC1)*B1(NC1))
C
      B1(NC1) = B1(NC1)*B1(NC1)
C
      RETURN
      END
C
C
      SUBROUTINE LKLIHD(Z,Y,N,P,Q,L)
      REAL N, L, LL
C
C           P = EXP(Z)/(1.+EXP(Z))
C           LL = Y*ALOG(P) + (N-Y)*ALOG(Q)
C           L = EXP(LL)
C
      EBND = 650.0
      IF(Z.GT.EBND) Z = EBND
      IF(Z.LT.-EBND) Z = -EBND
      E = EXP(Z)
      Q = 1./(1.+E)
      P = E*Q
C
      QLOG = -ALOG(1.+E)
      PLOG = Z + QLOG
      LL = Y*PLOG + (N-Y)*QLOG
C
      IF(LL.GT.EBND) LL = EBND
      IF(LL.LT.-EBND) LL = -EBND
      L = EXP(LL)
C
      RETURN
      END
```

```
      SUBROUTINE SINV(A,N,EPS,IER,DET)
      DIMENSION A(500)
C
C                 21 AUG 74
C          THIS PROGRAM FINDS AN APPROXIMATION TO THE INVERSE
C          OF A SYMETRIC, NOT NEC. POS. DEFINITE, MATRIX BY
C          ADDING TERMS TO THE DIAGONAL WHILE FACTORING.
C
C          FACTORIZE GIVEN MATRIX BY MEANS OF SUBROUTINE MFSD
C          A = TRANSPOSE(T) * T
C
      CALL MFSD(A,N,EPS,IER)
C
      IF(IER) 9,1,1
C
C          CALCULATE DETERMINANT OF GIVEN MATRIX
    1 J=0
      DET=1.
      DO 10 I=1,N
      J=J+I
   10 DET=DET*A(J)
      DET=DET*DET
C
C          INVERT UPPER TRIANGULAR MATRIX T
C          PREPARE INVERSION-LOOP
      IPIV = N*(N+1)/2
      IND = IPIV
C
C          INITIALIZE INVERSION-LOOP
      DO  6  I=1,N
      DIN = 1.0/A(IPIV)
      A(IPIV) = DIN
      MIN = N
      KEND = I - 1
      LANF = N - KEND
      IF(KEND) 5,5,2
    2 J = IND
C
C          INITIALIZE ROW-LOOP
      DO  4  K=1,KEND
      WORK = 0.0
      MIN = MIN - 1
      LHOR = IPIV
      LVER = J
      DO  3  L=LANF,MIN
C
C          START INNER LOOP
      LVER = LVER + 1
      LHOR = LHOR + L
    3 WORK = WORK + A(LVER) * A(LHOR)
C          END INNER LOOP
C
      A(J) = -WORK * DIN
    4 J = J - MIN
C          END ROW LOOP
C
```

```
    5 IPIV = IPIV - MIN
    6 IND = IND - 1
C         END OF INVERSION LOOP
C
C         CALCULATE INVERSE(A) BY MEANS OF INVERSE(T)
C         INVERSE(A) = INVERSE(T) * TRANSPOSE(INVERSE(T))
C         INITIALIZE MULTIPLICATION-LOOP
      DO 8 I=1,N
      IPIV = IPIV + I
      J = IPIV
C
C         INITIALIZE ROW-LOOP
      DO 8 K=I,N
      WORK = 0.0
      LHOR = J
C
C         START INNER LOOP
      DO 7 L=K,N
      LVER = LHOR + K - I
      WORK = WORK + A(LHOR) * A(LVER)
    7 LHOR = LHOR + L
C         END INNER LOOP
C
      A(J) = WORK
    8 J = J + K
C         END OF ROW- AND MULTIPLICATION-LOOP
C
    9 RETURN
      END
C
      SUBROUTINE MFSD(A, N, EPS, IER)
      DIMENSION A(500), D(32), DELTA(32)
C
C         TEST N
      IF(N-1) 22, 1, 1
    1 IER = 0
C
C         CALCULATE BETTA
      KPIV = 0
      KK = 1
      B1 = 0.
      B2 = 0.
      DO 4 K = 1, N
      KPIV = KPIV + K
      DO 4 J = 1, K
      IF (KK - KPIV) 3, 2, 22
    2 D(K) = A(KK)
      A(KK) = A(KK) + EPS
      IF(B1.LT.ABS(A(KK))) B1 = ABS(A(KK))
      GO TO 4
    3 IF(B2.LT.ABS(A(KK))) B2 = ABS(A(KK))
    4 KK = KK + 1
      B1 = SQRT(B1)*1.1
      B2 = SQRT(B2/N)
      B = B2
      IF(B1.GT.B2) B = B1
```

```
C
C           INITIALIZE DIAGONAL-LOOP
      KPIV = 0
      TOLSQ = 2.**(-24.)
C
      DO 17 K = 1, N
      TH = 0.
      KPIV = KPIV + K
      IND = KPIV
      LEND = K-1
C
C           START FACTORIZATION-LOOP OVER K-TH ROW
      DO 15 I = K, N
      DSUM = 0.
      IF(LEND) 22, 7, 5
C
C           START INNER LOOP
    5 DO 6 L = 1, LEND
      LANF = KPIV - L
      LIND = IND - L
    6 DSUM = DSUM + A(LANF)*A(LIND)
C           END INNER LOOP
C
C           TRANSFORM ELEMENT A(IND)
    7 DSUM = A(IND) - DSUM
C
      IF(I-K) 14, 8, 14
C
C           ADJUST DIAGONAL ELEMENT
    8 DELTA(K) = DSUM
      IF(DSUM-TOLSQ) 9, 13, 13
    9 IF(DSUM + TOLSQ) 10, 10, 11
   10 DSUM = -DSUM
      GO TO 12
   11 DSUM = TOLSQ
   12 IF(IER.LE.0) IER = K
C
C           COMPUTE PIVOT ELEMENT
   13 DPIV = SQRT(DSUM)
      A(KPIV) = DPIV
      GO TO 15
C
C           CALCULATE ROW TERMS AND MAX (R(K,IND))
   14 A(IND) = DSUM/DPIV
      IF(TH.LT.ABS(A(IND))) TH = ABS(A(IND))
C
   15 IND = IND + I
C
C           BOOST DIAGONAL ELEMENT FURTHER IF NECESSARY
      IF(TH.LE.0.) GO TO 17
      BTR = (B/TH)
      IF(BTR.GT.1.) GO TO 17
      IF(IER.LE.0) IER = K
      IND = KPIV + K
      A(KPIV) = A(KPIV)/BTR
      K1 = K+ 1
```

```
      DO 16 I = K1, N
      A(IND) = A(IND)*BTR
   16 IND = IND + I
   17 DELTA(K) = A(KPIV)*A(KPIV)-DELTA(K)+EPS
C            END OF DIAGONAL-LOOP
C
      IF(IER.LE.0) GO TO 19
      WRITE(61,18) (D(K),DELTA(K),K=1,N)
   18 FORMAT(/,6X,≠DIAGONAL TERMS AND ADJUSTMENTS ≠,
     1/,(6X,E12.5,6X,E12.5))
      GO TO 21
   19 WRITE(61,20) EPS
   20 FORMAT(/,2X,≠EPS =≠,F12.9)
   21 IER = 0
      RETURN
   22 IER = -1
      RETURN
      END
```

LUN3

U(I)	W(I)
2.4534070830E-01	4.6224366958E-01
7.3747372854E-01	2.8667550535E-01
1.2340762154E 00	1.0901720602E-01
1.7385377123E 00	2.4810520886E-02
2.2549740023E 00	3.2437733416E-03
2.7888060585E 00	2.2833863595E-04
3.3478545675E 00	7.8025564733E-06
3.9447640404E 00	1.0860693702E-07
4.6036824500E 00	4.3993409881E-10
5.3874808904E 00	2.2293936434E-13

```
      PROGRAM TABLE
      DIMENSION W(199),W1(199),W2(199),POW(31,5)
      A = .05
      DF = 8.
      EPS = .0001
      IDF = DF+.25
      REWIND 1
      READ(1,100)(W(I),I=1,IDF)
      WRITE(20,101)(W(I),I=1,IDF)
      SK = 0.
      DO 8 K=1,31
      POW(K,1) =  SK
    8 SK = SK+.05
      SNUL = 0.
      DO 18 I=2,4
      S = 0.
      DO 10 J=1,IDF
   10 W1(J) = W(J)/(1+SNUL*W(J))
      C = QTCR(A,DF,W1,EPS)
      DO 16 K=1,31
      DO 12 J=1,IDF
   12 W2(J) = W1(J)*(1+S*W(J))
      POW(K,I) = QT(C,DF,W2,EPS)
   16 S = S+.05
   18 SNUL = SNUL+.7
      DO 20 J=1,IDF
   20 W1(J) = 1.
      C = QTCR(A,DF,W1,EPS)
      S = 0.
      DO 24 K=1,31
      DO 22 J=1,IDF
   22 W2(J) = W1(J)*(1.+S*W(J))
      POW(K,5) = QT(C,DF,W2,EPS)
   24 S = S+.05
      WRITE(20,102)
      DO 26  I=1,31
   26 WRITE(20,103)(POW(I,J),J=1,5)
      CALL EXIT
  100 FORMAT(20F6.3)
  101 FORMAT(5X,≠W =≠,5(F6.3),/,1X,/,1X)
  102 FORMAT(16X,≠0.0≠,3X,≠0.7≠,3X,≠1.4≠,3X,≠WALD≠,/,1X)
  103 FORMAT(1X,F5.2,8X,4(F5.3,1X))
      END
C
C
```

```
      FUNCTION QTCR(A,DF,W,EPS)
      DIMENSION W(199)
      C = 0
      IDF = DF+.25
      DO 12 I=1,IDF
   12 C = C+W(I)
   14 F1 = QT(C,DF,W,EPS)
      E = A-F1
      IF(ABS(E).LT.10*EPS) GO TO 18
      IF(E.GT.0) GO TO 16
      C2 = C+100*EPS
      F2 = QT(C2,DF,W,EPS)
      DC = 100*EPS*E/(F2-F1)
      C = C+DC
      GO TO 14
   16 C2 = C-100*EPS
      F2 = QT(C2,DF,W,EPS)
      DC = -100*EPS*E/(F2-F1)
      C = C+DC
      GO TO 14
   18 QTCR = C
      RETURN
      END
C
C
```

```
      FUNCTION QT(XO,DF,W,EPS)
      DIMENSION W(199), E(999)
      DIMENSION F1(199), F2(199), H(999)
      N = DF+.25
C
      WMIN = W(1)
      WMAX = W(1)
      DO 2 I=1,N
      IF(W(I).LT.WMIN) WMIN = W(I)
      IF(W(I).GT.WMAX) WMAX = W(I)
    2 CONTINUE
      BTA = 2.*WMIN*WMAX/(WMIN+WMAX)
      X = XO/BTA
C
      P = 1.
      DO 4 I=1,N
      R = BTA/W(I)
      P = P*R
      F1(I) = 1.-R
    4 F2(I) = 1.
      EZ = SQRT(P)
C
      S2 = EZ*CHIA(X,DF)
      K = 1
      KEND = 0
    6 SI = DF+2.*K
C
      S1 = 0
      DO 8 I=1,N
      F2(I) = F2(I)*F1(I)
    8 S1 = S1 + F2(I)
      H(K) = S1
C
      S1 = H(K)*EZ
      IF(K.EQ.1) GO TO 12
      KM = K-1
      DO 10 I=1,KM
      M = K-I
   10 S1 = S1 + H(M)*E(I)
   12 E(K) = S1/(2.*K)
C
      S = E(K)*CHIA(X,SI)
      S2 = S2 + S
      K = K+1
      IF(K.GE.999) GO TO 14
      IF(ABS(S).LE.EPS) KEND = KEND+1
      IF(KEND.LT.3) GO TO 6
C
   14 QT = 1.-S2
      RETURN
      END
C
```

```
      FUNCTION CHIA(X,DF)
      IF(X.LE.0) GO TO 3
      D = DF/2.
      S = 0
      PRX = X/2.
      PR = 1
      J = 0
    1 DIV = 1/(D+J)
      PR = PR * PRX * DIV
      S = S+PR
      J = J+1
      IF(J.GE.999) GO TO 2
      IF(PR.GT..0000000001) GO TO 1
    2 CM = EXP(-X/2.)*((X/2.)**(D-1))
      CHIA = (CM*S)/GAM(D)
      RETURN
    3 CHIA = 0.
      RETURN
      END
C
C
      FUNCTION GAM(DD)
C
C          DD=N/2 FOR SOME INTEGER N, 1 .LE. DD .LE. 99
C
      D = DD
      GME = GMO = 1.
      I = 0
      IF(D.GT.99) GO TO 4
      IF(D.LT.0.25) GO TO 4
    1 CONTINUE
      IF(D.LT.0.75) GO TO 2
      IF(D.LT.1.25) GO TO 3
      GME = (I+1.0)*GME
      GMO = (I+0.5)*GMO
      D = D-1.
      I = I+1
      GO TO 1
    2 GAM = GMO*1.77246
      RETURN
    3 GAM = GME
      RETURN
    4 WRITE(61,5)
    5 FORMAT(1X,≠GAM ARGUMENT OUTSIDE BOUNDS≠)
      CALL EXIT
      END
```