8129108

TEOH, KOK WAH

CONTRIBUTIONS TO THE ASYMPTOTIC THEORY OF ESTIMATION AND HYPOTHESIS TESTING WHEN THE MODEL IS INCORRECT

The Ohio State University PH.D. 1981

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

CONTRIBUTIONS TO THE ASYMPTOTIC THEORY OF ESTIMATION AND HYPOTHESIS TESTING WHEN THE MODEL IS INCORRECT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

by

Kok Wah Teoh, B.S., M.S.

* * * * *

The Ohio State University

1981

Reading Committee:

Ramesh C. Srivastava

D. Ransom Whitney

Jagdish S. Rustagi

Approved by

R. C. Srivastava

Advisor
Department of Statistics

ACKNOWLEDGMENTS

I wish to express my deep appreciation to Professor Ramesh Srivastava for his guidance, advice and patient encouragement during the writing of this thesis, and also throughout these past four years at the Department of Statistics, OSU.

I also wish to thank Professors Jagdish Rustagi and Ransom Whitney for their helpful comments on this thesis.

VITA

May 1, 1952	Born - Alor Star, Malaysia
1976	B.S. (Honours) - University of Malaya, Kuala Lumpur, Malaysia
1978	M.S. (Mathematics) - University of Malaya, Kuala Lumpur, Malaysia
1977-1981	Teaching Associate - Department of Statistics, The Ohio State University, Columbus, Ohio
1979	M.S. (Statistics) - The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Mathematical Statistics

Studies in Nonparametric Statistics - Professors Michael A. Fligner and Douglas A. Wolfe

Studies in Analysis of Variance and Experimental Design - Professor C. Radhakrishna Rao

Studies in Multivariate Analysis - Professor Ramesh C. Srivastava

Studies in Decision Theory - Professor Jason Hsu

Studies in Statistical Inference - Professor Michael A. Fligner

Studies in Probability Theory - Professor Louis Sucheston

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The investigation of the asymptotic properties of the maximum likelihood estimator (m.l.e.) when the model is incorrect was first begun by Huber in 1965. More recently, Foutz and Srivastava (1974) established the asymptotic distribution of the likelihood ratio test statistic when the model is incorrect. In this work, we strengthen and expand on some of the large-sample results of the m.l.e. and the likelihood ratio test statistic under the incorrect model. In addition, we will also study the asymptotic properties of the Rao and the Wald statistics, and compare the performance of these three test statistics.

In Section 1.1, we briefly outline the development of research in this area. The consistency and asymptotic normality of the m.l.e. are reviewed in Section 1.2, the asymptotic distribution of the likelihood ratio test statistic under model misspecification is presented in Section 1.3 and a method of stochastic comparison of tests used in Foutz and Srivastava (1977) is defined in Section 1.4. Finally, we motivate and indicate the direction of this research in Section 1.5.

1.1 Preliminaries

The method of maximum likelihood estimation, first proposed by Fisher in 1921, has been for years one of the most important tools of statistical inference. The controversies surrounding it, especially with respect to certain optimality properties that were ascribed to it, have plagued the statistical community for quite some time and it is only in recent years that most of these are resolved satisfactorily with some measure of completeness and finality.

Associated with this method of estimation is the general hypothesis testing procedure called the likelihood ratio test, first proposed by Neyman and Pearson. As with the maximum likelihood estimator, under certain conditions, the likelihood ratio test possesses some nice optimality properties. In addition, there are two other important large-sample test statistics due to Rao and Wald that also utilize the m.l.e.

A very crucial assumption implicit in the proofs of all these optimality results is the assumption that the probability model used to construct the m.l.e. and the related test statistics has been correctly specified. The question of model misspecification assumes great importance in the case of the m.l.e. due to its lack of robustness in many aspects. Even its status as a sufficient statistic, when it applies, relies heavily on the assumption of correct model specification. It is thus of great value to study the behavior of the m.l.e. and related statistics when the model is incorrect.

Let X_1, X_2, ... be a sequence of independent and identically distributed (i.i.d.) random variables defined on the probability space (χ, B, P). Let F(x) be the distribution function (d.f.) associated with P; i.e., $F(x) = P(X_1 \leq x)$, $-\infty < x < \infty$. In most problems of statistical inference, it is assumed that there exists a density g(x) of F(x) with respect to some σ-finite measure and that g(x) belongs to a parametric family of density functions, $\{f(x, \theta), \theta \varepsilon \Theta\}$, where the index set Θ is an open subset of the k-dimensional Euclidean space E^k. This parametric family is called the model and it is assumed that $g(x) = f(x, \theta_0)$ for some $\theta_0 \varepsilon \Theta$. The estimation problem is then to find an estimator (preferably one that possesses nice properties such as unbiasedness, minimum variance, etc.) and a confidence region for θ_0. The corresponding hypothesis testing problem is to determine if θ_0 belongs to a subset Θ_0 of Θ.

The model is correctly specified, or equivalently, the model is correct, if g(x) belongs to the specified model, i.e.,

$$g(x) = f(x, \theta) \qquad \text{for some } \theta \varepsilon \Theta .$$

Otherwise, the model is misspecified (or equivalently, incorrect). However, we will use the term 'misspecified' and 'incorrect' loosely to include also the case when the model is correctly specified.

1.2 Asymptotic Properties of the Maximum Likelihood Estimator When the Model is Incorrect

Huber (1965) was the first to prove the consistency and asymptotic normality of the m.l.e. when the model is incorrect. Since then, other authors such as Foutz and Srivastava (1974), and White (1980) have shown the consistency and asymptotic normality of the m.l.e. under various regularity conditions different from those of Huber.

In this section, we quote the results contained in Foutz and Srivastava (1974). Let the likelihood function of $X_1, X_2, \ldots, X_n$ be defined as

$$L_n(\theta) = \Pi_{i=1}^{n} f(X_i, \theta) .$$

Denote by $\frac{\partial}{\partial\theta} L_n(\theta)$ the k x 1 vector $(\frac{\partial}{\partial\theta_1} L_n(\theta), \ldots, \frac{\partial}{\partial\theta_k} L_n(\theta))'$. The m.l.e. $\hat{\theta}_n$ of θ is a solution to the likelihood equation

$$\frac{\partial}{\partial\theta} L_n(\theta) = 0.$$

In their proof of the consistency of the m.l.e., Foutz and Srivastava (1974) assumed the following regularity conditions:

Let $E(\cdot)$ denote the expectation taken w.r.t. $F(x)$.

A1. There exists a $\theta^* \varepsilon \Theta$ such that

$$E[\log f(X, \theta^*)] = \sup_{\Theta} E[\log f(X, \theta)].$$

A2. The partial derivatives

$$\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} E[\log f(X, \theta)]$$

ℓ, m = 1, 2, ..., k exist and are continuous in an open neighborhood Θ^* of θ^*.

A3. The matrix

$$\Lambda(\theta) = \left(E[\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X, \theta)] \right)_{k \times k}$$

ℓ, m = 1, 2, ..., k is nonsingular for $\theta \varepsilon \Theta^*$.

A4. For every $\theta \varepsilon \Theta^*$,

$$\frac{\partial}{\partial\theta_\ell} E[\log f(X, \theta)] = E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta)]$$

ℓ = 1, 2, ..., k.

<u>Theorem 1.2.1 (Foutz and Srivastava)</u>: Let X_1, X_2, ... be a sequence of i.i.d. random variables having common d.f. F(x). Let the family of density functions, {f(x, θ), θεΘ}, be an assumed model where no f(x, θ) in the model need be a density for F(x). Then with θ* as in condition A1, conditions A2, A3 and A4 insure the existence of a sequence of solutions of the likelihood equation

$$\frac{\partial}{\partial\theta} \log L_n(\theta) = 0, \; n = 1, 2, \ldots$$

which converges almost surely to θ^* as $n \to \infty$.

To establish the asymptotic normality of the m.l.e., Foutz and Srivastava (1974) assume another set of regularity conditions which essentially build on those needed in the proof of consistency. These regularity conditions are:

B1. The m.l.e. for θ in the model $\{f(x, \theta), \theta\varepsilon\Theta\}$ converges in probability to the constant θ^*, uniquely satisfying the condition

$$E[\log f(X, \theta^*)] = \sup_{\Theta} E[\log f(X, \theta)] \; .$$

B2. $E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta)]$ and $E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta) \frac{\partial}{\partial\theta_m} \log f(X, \theta)]$ are finite in a neighborhood Θ^* of θ^* for ℓ, $m = 1, 2, \ldots, k$.

B3. $E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta)]$ is finite in Θ^*, ℓ, $m = 1, 2, \ldots, k$.

B4. $E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta)] = \frac{\partial}{\partial\theta_\ell} E[\log f(X, \theta)]$, $\ell = 1, 2, \ldots, k$.

B5. $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta) \xrightarrow{p} E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta)]$ uniformly in $\theta\varepsilon\Theta^*$, ℓ, $m = 1, 2, \ldots, k$, where $\xrightarrow{p}$ denote convergence in probability.

B6. The matrix

$$\Lambda(\theta) = \left(E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta)] \right)_{k \times k}$$

ℓ, $m = 1, 2, \ldots, k$ is nonsingular for $\theta\varepsilon\Theta^*$.

Also define the 'pseudo' information matrix

$$C(\theta^*) = \left(E[\frac{\partial}{\partial\theta_\ell} \log f(X,\theta^*) \frac{\partial}{\partial\theta_m} \log f(X, \theta^*)] \right)_{k \times k}$$

$\ell, m = 1, 2, \ldots, k.$

<u>Theorem 1.2.2 (Foutz and Srivastava)</u>: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having common d.f. F(x). Let the assumed model, $\{f(x, \theta), \theta\varepsilon\Theta\}$, satisfy regularity conditions B1, B2, ..., B6 w.r.t. F(x). Then, with $\hat{\theta}_n$ the m.l.e. for θ in the assumed model, $\sqrt{n}(\hat{\theta}_n - \theta^*)$ is asymptotically normally distributed as $n \to \infty$ with mean vector 0 and covariance matrix

$$\Lambda^{-1}(\theta^*)C(\theta^*)\Lambda^{-1}(\theta^*).$$

1.3 The Asymptotic Distribution of the Likelihood Ratio Statistic When the Model is Incorrect

Let Θ_0 be a subset of Θ defined by

$$\Theta_0 = \{\theta | \theta\varepsilon\Theta,\ \theta_j = \theta_{0j},\ j = 1, 2, \ldots, r\}$$

where $1 \leq r \leq k$ and $\delta_0 = (\theta_{01}, \ldots, \theta_{0r})'$ is a r x 1 vector of specified constants. Denote $\theta' = (\delta', \gamma')$ where $\delta = (\theta_1, \ldots, \theta_r)'$ and $\gamma = (\theta_{r+1}, \ldots, \theta_k)'$. The hypothesis testing problem of interest is to test the hypothesis H_0: $\theta\varepsilon\Theta_0$ against the alternative H_1: $\theta\varepsilon\Theta - \Theta_0$.

The likelihood ratio is defined as

$$\lambda_n = \frac{\sup\{L_n(\theta):\ \theta\varepsilon\Theta_0\}}{\sup\{L_n(\theta):\ \theta\varepsilon\Theta\}} = \frac{L_n(\tilde{\theta}_n)}{L_n(\hat{\theta}_n)}$$

where $\tilde{\theta}_n(\hat{\theta}_n)$ is the restricted (unrestricted) m.l.e. of θ over $\Theta_o(\Theta)$. The likelihood ratio test statistic proposed by Neyman and Pearson is defined as

$$T_n = -2\log \lambda_n.$$

It is well-known that when the model is correct, i.e.,

$$g(x) = f(x, \theta_0) \text{ for some } \theta_0\varepsilon\Theta ,$$

the likelihood ratio statistic T_n has an asymptotic chi-squared distribution with r degrees of freedom under the null hypothesis. Further $T_n \xrightarrow{P} \infty$ as $n \to \infty$ under any alternative $\theta\varepsilon\Theta-\Theta_0$, so that the test is consistent.

In their 1978 paper, Foutz and Srivastava derived the asymptotic distribution of the likelihood ratio statistic T_n when the model is incorrect. We quote below the regularity conditions used to prove their results.

C1. Assume condition A1.

C2. Assume the k x k matrix $\Lambda(\theta^*)$ with (ℓ,m)-th element

$$E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta^*)], \ \ell, m = 1, 2, \ldots, k$$

exists and is nonsingular. Also assume the sequence $\{\hat{\theta}_n\}$ converges in probability to θ^* and assume $\sqrt{n}(\hat{\theta}_n - \theta^*)$ is asymptotically normal with mean 0 and covariance matrix

$$\Lambda^{-1}(\theta^*)C(\theta^*)\Lambda^{-1}(\theta^*) \ .$$

C3. Assume $E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta^*)] = 0$, $\ell = 1, 2, \ldots, k$ and

$$\frac{\partial}{\partial\theta_\ell} E[\frac{\partial}{\partial\theta_m} \log f(X, \theta^*) = E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta^*)] \ ,$$

$$\ell, m = 1, \ldots, k.$$

C4. Assume

$$\sup_{\Theta^*} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta) - E[\frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X, \theta)] \right| \xrightarrow{P} 0$$

for some neighborhood Θ^* about θ^*.

Let $\xrightarrow{D}$ denote convergence in distribution.

Theorem 1.3.1 (Foutz and Srivastava): Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common d.f. F(x). Let the assumed model, $\{f(x, \theta), \theta\varepsilon\Theta\}$, satisfy regularity conditions C1 - C4. Denote by M the upper r x r diagonal block of the matrix $\Lambda^{-1}(\theta^*)C(\theta^*)\Lambda^{-1}(\theta^*)$ of conditions C2. Partition $\Lambda(\theta^*)$ in a form having upper r x r diagonal block Λ_{11}:

$$\Lambda(\theta^*) = \begin{bmatrix} \Lambda_{11} & \Lambda_{21}' \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

and write $W = -\Lambda_{11} + \Lambda_{21}' \Lambda_{22}^{-1} \Lambda_{21}$. If $\theta^*\varepsilon\Theta_0$, then T_n is asymptotically distributed as a linear combination of independent chi-squared random variables:

$$T_n \xrightarrow{D} \sum_{i=1}^{r} c_i \chi_i^2$$

where $\chi_1^2, \chi_2^2, \ldots, \chi_r^2$ are independent chi-squared random variables with 1 degree of freedom and $c_1 \geq c_2 \geq \ldots \geq c_r$ are the eigenvalues of the matrix MW.

Foutz and Srivastava (1978) also indicated the 'non-null' behavior of T_n in

Theorem 1.3.2 (Foutz and Srivastava): Assume the existence of a unique θ_0^* that maximizes $E[\log f(X, \theta)]$ over Θ_0, and assume the model, $\{f(x, \theta), \theta\varepsilon\Theta\}$, satisfy regularity conditions C1 - C4.

Assume that for some open neighborhoods Θ^* and Θ_0^* about θ^* and θ_o^*

$$\sup_{\Theta^* \cup \Theta_0^*} \left| \frac{1}{n} \sum_{i=1}^{n} \log f(X_i, \theta) - E[\log f(X, \theta)] \right| \xrightarrow{p} 0 .$$

If $\theta^* \notin \Theta_0$, then

$$T_n/n \xrightarrow{p} 2\{E[\log f(X, \theta^*) - E[\log f(X, \theta_0^*)]\} .$$

Since the probability limit is a positive constant, this shows that $T_n \to +\infty$ as $n \to \infty$ for $\theta^* \notin \Theta_0$ and hence, establishes the 'consistency' of the likelihood ratio test even when the model used to construct the m.l.e. is incorrect.

1.4 Stochastic Comparison of the Performance of the Likelihood Ratio Test Under Model Misspecification

In another paper, Foutz and Srivastava (1977), a method for examining the performance of the likelihood ratio test when the model is incorrect is proposed. The resulting measures of asymptotic efficiency are based on the concept of Bahadur efficiency.

It is shown in the above paper that when the model is incorrect, the likelihood ratio statistic is a 'standard sequence' and an expression for the 'approximate slope' is derived. The ratio of the approximate slopes under the incorrect and correct models (as $\theta \to \theta_0$, if need be) then affords a measure of the (local, if $\theta \to \theta_0$) asymptotic relative efficiency of the likelihood ratio test for various departures from the assumed model.

These concepts will be more clearly defined in Chapter 4.

1.5 Direction of Research

In terms of practical applications such as the construction of confidence intervals or confidence regions and hypothesis testing, it is not enough just to establish the asymptotic normality of the m.l.e., $\hat{\theta}_n$, when the model is incorrect. In ordinary situations under correct model specification, authors such as Rao (1963) and Wolfowitz (1965) have pointed out the need to strengthen asymptotic normality to the stronger property of uniform asymptotic normality. In view of this, we shall establish appropriate regularity conditions in order for uniform asymptotic normality of the m.l.e. to hold under model misspecification.

Also, it is of interest to study the large-sample performance of the likelihood ratio test against so-called 'local alternatives'. For this purpose, we need the asymptotic distribution of the likelihood ratio test statistic under local alternatives when the model is incorrect. The above two results are contained in Chapter 2.

Next, we will focus on the test statistics due to Rao (1947) and Wald (1943). Let $E_\theta(\cdot)$ denote expectation taken w.r.t. the density $f(x, \theta)$. Define the 'quasi' information matrix

$$I(\theta) = \left\{E_\theta\left[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta) \frac{\partial}{\partial\theta_m} \log f(X, \theta)\right]\right\}_{k \times k}$$

$$\ell, m = 1, 2, \ldots, k$$

and the k x 1 vector

$$V_\theta = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta_1} \log f(X_i, \theta), \ldots, \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta_k} \log f(X_i, \theta)\right)' .$$

Partition $\hat{\theta}_n' = (\hat{\delta}_n', \hat{\gamma}_n')$, $\tilde{\theta}_n' = (\delta_0', \tilde{\gamma}_n')$ where $\hat{\delta}_n$ and δ_0 are r x 1 vectors. The Rao statistic is defined as

$$R_n = V_{\tilde{\theta}_n}' \, I^{-1}(\tilde{\theta}_n) \, V_{\tilde{\theta}_n} ,$$

and the Wald statistic is defined as

$$W_n = n(\hat{\delta}_n - \delta_0)'[(I_r \quad 0) \, I^{-1}(\hat{\theta}_n) \binom{I_r}{0'}]^{-1} (\hat{\delta}_n - \delta_0),$$

where I_r is the r x r identity matrix and 0 is a rx(k-r) matrix of zeros. (Note that the matrix $(I_r \quad 0) \, I^{-1}(\hat{\theta}_n)\binom{I_r}{0'}$ is just the upper r x r diagonal block of $I^{-1}(\hat{\theta}_n)$.) These two statistics provide general tests of the hypothesis H_0 of Section 1.3.

Both the Rao and the Wald statistics utilize the m.l.e. and when the model is correct, it is well-known that they have the same asymptotic distribution as the likelihood ratio test statistic, viz., chi-squared distribution with r degrees of freedom. In Chapter 3, we derive the asymptotic distribution of the Rao and Wald statistics when the model is incorrect, first, under the 'null' hypothesis and, second, under a sequence of 'local' alternatives'.

Under correct model specification, due to the above distributional equivalence, it has been traditionally difficult to compare and judge the performance of all the three test statistics considered. In Chapter 4, using the asymptotic results of the preceding chapters, we shall attempt to compare these test statistics. Our approach follows that of Foutz and Srivastava (1977) and is based on a modification of the concept of Bahadur efficiency.

The last chapter is devoted to a comparison of the likelihood ratio statistic vis-a-vis the Rao statistic using an optimality criterion introduced by Bahadur for the situation when the model is correctly specified. We will show that the Rao statistic is not optimal according to this criterion. Further, we will also examine the special cases of the 1-parameter exponential model and the normal model which is a member of the 2-parameter exponential family.

CHAPTER 2

UNIFORM ASYMPTOTIC NORMALITY OF MAXIMUM LIKELIHOOD ESTIMATOR AND ASYMPTOTIC DISTRIBUTION OF LIKELIHOOD RATIO STATISTIC UNDER LOCAL ALTERNATIVES WHEN THE MODEL IS INCORRECT

To facilitate the application of the results contained in Theorem 1.2.2 to the construction of confidence regions and hypothesis testing, we need to strengthen it to one of uniform asymptotic normality under the incorrect model.

In anticipation of the need to study the local performance, under model misspecification, of the likelihood ratio test of the hypothesis H_0 against alternatives close to the null-hypothesis value, we should also make available the asymptotic distribution of the likelihood ratio test statistic under model misspecification against a sequence of local alternatives of the form

$$\theta^* = \theta_0 + \frac{\Delta}{\sqrt{n}} \quad , n = 1, 2, \ldots .$$

These two results are established in the following sections.

2.1 Preliminaries

In the usual setting of the estimation problem when the model is correct, it is customary to first establish the uniform asymptotic normality for the standardized m.l.e. $\sqrt{n}\,(\hat{\theta}_n - \theta)$. Consider the unidimensional case where k = 1. For a prespecified confidence coefficient of $(1 - \alpha)$, the corresponding cutoff $z_{\alpha/2}$ is determined from standard normal tables and the $(1 - \alpha)100\%$ confidence interval for θ is then obtained by inverting the inequalities

$$\frac{-Z_{\alpha/2}}{\sqrt{I(\theta)}} \leq \sqrt{n}(\hat{\theta}_n - \theta) \leq \frac{Z_{\alpha/2}}{\sqrt{I(\theta)}}$$

to yield the confidence interval

$$\hat{\theta}_n - \frac{Z_{\alpha/2}}{\sqrt{nI(\theta)}} \leq \theta \leq \hat{\theta}_n + \frac{Z_{\alpha/2}}{\sqrt{nI(\theta)}}$$

where $I(\theta)$ is the Fisher information per observation. It should be noted that if uniformity in θ does not hold, then the actual confidence level attained varies for different θ values. Or to put it in another way, different values of $n = n(\theta)$ are required for each value of θ in order to attain the prespecified confidence level of $(1 - \alpha)$.

Similarly, when the model is incorrect, i.e.

$$f(x, \theta) \neq g(x), \qquad \text{for all } \theta \varepsilon \Theta ,$$

it could still be useful to know which member of the model $\{P_\theta, \theta\varepsilon\Theta\}$ provide the closest approximation to the true probability measure P. This naturally leads to the question of which f(x, θ) is 'closest' in some sense to g(x). The following argument justifies θ* (as defined in condition A1) as the proper value of θ that makes f(x, θ) 'closest' to g(x) in an information theoretic sense.

2.2 Uniform Asymptotic Normality of the Maximum Likelihood Estimator under Model Misspecification

In many realistic situations, a statistician cannot be absolutely certain that the model he has chosen to use to construct the m.l.e. is the correct one. Even if he had carried out a standard goodness-of-fit test to determine the correctness of his model, the acceptance of his null-hypothesis merely asserts that there is insufficient evidence to reject the model based on the observations that he had obtained. At best, this could be interpreted as saying that his model is a close approximation to the true underlying distribution. It would be grossly over-optimistic to claim, upon acceptance of the null-hypothesis, that the chosen model is 'the' correct one. This seemingly anomalous interpretation of the results stems from the striking observation that whenever a goodness-of-fit test results in the acceptance of the null-hypothesis that the observations come from the family $\{P_\theta, \theta\varepsilon\Theta\}$, the very same test (based on the same data) will also accept the hypothesis that the observations come from a bigger family, say $\{P_{\theta,\gamma}, (\theta, \gamma)\varepsilon\Theta x\Gamma\}$, containing $\{P_\theta, \theta\varepsilon\Theta\}$. In typical situations, this bigger family could take the form of a 'generalized' version of the original model and we wish to work with the 'limited' model $\{P_\theta, \theta\varepsilon\Theta\}$ because it may be more convenient and mathematically tractable than the generalized version. An example of this arises when $\{P_\theta, \theta\varepsilon\Theta\}$ is the family of Poisson distributions with mean $\theta > 0$ and the bigger family is the hyper-Poisson distributions with parameters $\theta > 0$ and $0 < \lambda < \infty$ defined by the density, $f(x, \theta,\lambda) = K(\theta,\lambda)\ \theta^x\Gamma(\lambda)/\Gamma(x+\lambda)$, $x = 0, 1, \ldots$, where $K(\theta, \lambda)$ is a summation constant depending only on θ and λ. Another example occurs if we take $\{P_\theta, \theta\varepsilon\Theta\}$ to be the family of normal distributions with mean μ and variance σ^2 (i.e. $\theta = (\mu, \sigma^2)'$) and the bigger family represents the contaminated normal distributions with means μ_1, μ_2 and variances σ_1^2, σ_2^2 and a mixture proportion $0 \leq p \leq 1$.

From the preceding discussion, it is thus clear that it is very useful to know which member of the chosen model $\{P_\theta, \theta\varepsilon\Theta\}$ 'best' approximates the true underlying probability measure P. We shall justify below why $P_{\theta*}$ can claim to be closest to P.

Let us first define the Kullback-Leibler information measure $I_g(f)$ of $f(x, \theta)$ w.r.t. $g(x)$ as

$$\begin{aligned} I_g(f) &= \int \log(g/f)dP \\ &= E(\log g) - E(\log f). \end{aligned}$$

It is well-known that $0 \leq I_g(f) \leq \infty$ for all density f and $I_g(f) = 0$ iff $g = f$ a.s. P. Hence, we can regard $I_g(f)$ as a measure of the distance or 'similarity' between the probability measures P_θ and P. By idenfifying the θ^* that maximizes the quantity $E(\log f(X, \theta))$, we are therefore seeking the member $P_{\theta*}$ of the model $\{P_\theta, \theta\varepsilon\Theta\}$ that is closest or most 'similar' to P in terms of the Kullback-Leibler information measure $I_g(f)$.

To implement the search for θ^*, we need an estimation procedure that will lead to some kind of an approximate confidence interval (region) for θ^*. Having knowledge of the results of the previous chapter, an obvious approach therefore is to use the maximum likelihood method. As pointed out earlier, the asymptotic normality of the m.l.e. under the incorrect model has already been established by several authors under various regularity conditions. However, for purposes of constructing confidence regions for θ^*, we need the stronger condition of uniform asymptotic normality (over compact sets of θ^*) to hold as well. The following theorem gives a set of regularity conditions under which uniform asymptotic normality holds.

We shall denote the uniform weak convergence of a sequence of probability measures by $\Rightarrow_u$. Let C be the space of all real-valued bounded uniformly continuous functions. Let P and $\{P_\theta^{(n)}, \theta \varepsilon \Theta\}_{n=1}^\infty$ be probability measures. Then by definition,

$$P_\theta^{(n)} \Rightarrow_u P \quad \text{iff} \quad \int f \, dP_\theta^{(n)} \to \int f \, dP$$

as $n \to \infty$ uniformly in θ for all $f \varepsilon C$.

We will first establish some lemmas that will be needed later.

Lemma 2.2.1: If E $|X|^k$ and E $|Y|^k$ are both finite (k is a positive integer), then E $|X + Y|^k$ is also finite.

Proof. First, we will show that

$$|X + Y|^k \leq 2^{k-1}(|X|^k + |Y|^k) .$$

Indeed, it is true for k = 1. Since $|X + Y|^k \leq ||X| + |Y||^k$, without loss of generality, we can assume that X and Y are positive. Assume that the above is true for k-1. Then

$$\begin{aligned}(X + Y)^k &= (X + Y)^{k-1}(X + Y) \\ &\leq 2^{k-2}(X^{k-1} + Y^{k-1})(X + Y) \\ &= 2^{k-2}(X^k + Y^k + XY^{k-1} + YX^{k-1}) .\end{aligned}$$

It is enough to show

$$XY^{k-1} + YX^{k-1} \leq X^k + Y^k .$$

Case 1. X = Y. The above is trivially true in this case.

Case 2. It is enough to consider the case when X < Y. (The other case when X > Y follows by symmetry.)

$$\begin{aligned}XY^{k-1} + YX^{k-1} &= XY^{k-1} + (X + Y - X)Y^{k-1} \\ &\leq XY^{k-1} + X^k + (Y - X)Y^{k-1} \\ &= X^k + Y^k. \quad //\end{aligned}$$

Finally, we have

$$\begin{aligned}E\,|X + Y|^k &\leq 2^{k-1}E(\,|X|^k + |Y|^k) \\ &= 2^{k-1}(E\,|X|^k + E|Y|^k) < \infty. \quad //\end{aligned}$$

Lemma 2.2.2. If $\hat{\theta}_n \xrightarrow{p} \theta^*$ uniformly in θ^* and f is continuous at θ^*, then

$$f(\hat{\theta}_n) \xrightarrow{p} f(\theta^*) \qquad \text{uniformly in } \theta^* .$$

Proof. Given $\varepsilon > 0$ and $\delta > 0$, choose $\delta > 0$ such that $|f(\theta) - f(\theta^*)| < \varepsilon$ whenever $|\theta - \theta^*| < \delta$. For this ε, consider the following.

$$\begin{aligned}
&P\{|f(\hat{\theta}_n) - f(\theta^*)| > \varepsilon \} \\
&= P\{|f(\hat{\theta}_n) - f(\theta^*)| > \varepsilon, |\hat{\theta}_n - \theta^*| < \gamma\} \\
&\quad + P\{|f(\hat{\theta}_n) - f(\theta^*)| > \varepsilon, | \hat{\theta}_n - \theta^*| \geq \gamma\} \\
&\leq P\{|f(\hat{\theta}_n) - f(\theta^*)| > \varepsilon, |\hat{\theta}_n - \theta^*| < \gamma\} \\
&\quad + P\{|\hat{\theta}_n - \theta^*| \geq \gamma\} .
\end{aligned}$$

The first term equals zero by the continuity assumption made on f. By the uniform convergence of $\hat{\theta}_n$, we can choose an N such that for all $n > N$,

$$P(|\hat{\theta}_n - \theta^*| \geq \gamma) < \delta$$

and so

$$P(|f(\hat{\theta}_n) - f(\theta^*)| > \varepsilon) < \delta$$

for all $n > N$ where N does not depend on θ. //

The following notations are needed for Lemma 2.2.3. Let f be a real-valued measurable function on E^k. Define

$$\omega_f(A) = \sup\{|f(x) - f(y)| \;:\; x, y \in A\}$$

where $A \subset E^k$, and

$$\omega_f(x : \varepsilon) = \omega_f(B(x : \varepsilon)), \; x \in E^k, \; \varepsilon > 0$$

and $B(x : \varepsilon)$ is the open ball centered at x and of radius ε. Also define

$$\overline{\omega}_f (\varepsilon : \mu) = \int \omega_f(x : \varepsilon) \mu (dx)$$

where μ is a probability measure, and

$$\omega_f^* (\varepsilon : \mu) = \sup\{ \overline{\omega}_{f_y} (\varepsilon : \mu) : y \varepsilon E^k\}$$

where $f_y(x) = f(x+y)$ is the translate of f by y.

The following lemma is taken from Bhattacharya and Rao (1976) [Theorem 13.2, p. 113].

Lemma 2.2.3: Let $X_1, X_2, \ldots, X_n$ be n i.i.d. random vectors with values in E^k satisfying $E(X_1) = 0$, $Var(X_1) = I$ and $\rho_4 = E ||X_1||^4 < \infty$ where $||u|| = (\sum_{i=1}^k u_i^2)^{1/2}$ is the Euclidean norm. Let Q_n denote the distribution of $(X_1 + X_2 + \ldots + X_n)/\sqrt{n}$ and let Φ be the standard normal distribution in E^k. Then for every real, bounded Borel-measurable function f on E^k,

$$\begin{aligned} &\left| \int f \, d\, Q_n - \int f \, d\, \Phi \right| \\ &\leq \omega_f(E^k)\, a(k)\, \rho_4/\sqrt{n} \\ &\quad + \frac{4}{3} \cdot \omega_f^*(2^{7/2} k^{4/3} \rho_3/(\pi^{1/3}\sqrt{n}) : \Phi) , \end{aligned}$$

where $\rho_3 = E ||X_1||^3$.

We are now in a position to prove Theorem 2.2.1. The following regularity conditions are assumed to hold.

D1. For almost all x, the second order partial derivatives $\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(x, \theta)$, $\ell, m = 1, 2, \ldots, k$ exist and are continuous in θ.

D2. The m.l.e. for θ in the model $\{f(x, \theta), \theta\varepsilon\Theta\}$ converges in probability uniformly in $\theta^*\varepsilon\Theta$ to the constant θ^* uniquely satisfying the condition

$$E(\log f(X, \theta^*)) = \sup_{\Theta} \{E[\log f\ (x, \theta)]\} .$$

(This is the same as condition B1.)

D3. $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X_i, \theta) \xrightarrow{p} E[\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X, \theta)]$

uniformly in a neighborhood Θ^* of θ^*, $\ell, m = 1, 2, \ldots, k$. (This is the same as condition B5.)

D4. The matrix

$$\Lambda(\theta) = \left(E[\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X, \theta)] \right)_{k \times k}$$

$\ell, m = 1, 2, \ldots, k$ is nonsingular for $\theta\varepsilon\Theta^*$. (This is the same as condition B6.)

D5. Assume $E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta^*)]^4$, $\ell = 1, 2, \ldots, k$ exist and are finite.

D6. Assume the 'pseudo' information matrix $C(\theta^*)$ is positive definite where

$$C(\theta^*) = \left[E[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta^*) \frac{\partial}{\partial\theta_m} \log f(X, \theta^*)] \right]_{k \times k}$$

$\ell, m = 1, 2, \ldots, k.$

Theorem 2.2.1. Under regularity conditions D1 - D6,

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \Rightarrow_u N(0, \Lambda^{-1}(\theta^*)C(\theta^*)\Lambda^{-1}(\theta^*)) .$$

Proof. Let $\phi_\ell = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta^*)$, $\ell = 1, 2, \ldots, k$ and write $\phi = (\phi_1, \phi_2, \ldots, \phi_k)'$. For sufficienty large values of n, D1 and D2 allow us to expand ϕ by Taylor series about $\hat{\theta}_n = (\hat{\theta}_{n1}, \hat{\theta}_{n2}, \ldots, \hat{\theta}_{nk})'$.

$$\phi_\ell = \phi_\ell \mid_{\hat{\theta}_n} + \sum_{m=1}^{k} (\theta_m^* - \hat{\theta}_{nm}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta_1^+)$$

for some $\theta_\ell^+ = \hat{\theta}_n + \lambda_\ell(\theta^* - \hat{\theta}_n)$, $0 \leq \lambda_\ell \leq 1$ and $\ell = 1, 2, \ldots, k$. By the maximizing property of the m.l.e.,

$$\phi_\ell|_{\hat{\theta}_n} = 0, \ \ell = 1, 2, \ldots, k.$$

So we can write

$$\phi_\ell = -\sum_{m=1}^{k} \sqrt{n}(\hat{\theta}_{nm} - \theta^*) \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta_\ell^+) .$$

In matrix notation, this reduces to

$$\phi = \Lambda(\theta^+) \sqrt{n}(\hat{\theta}_n - \theta^*),$$

where

$$\hat{\Lambda}(\theta^+) = \left[-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell \partial\theta} \log f(X_i, \theta_\ell^+) \right]_{k \times k}$$

ℓ, m = 1, 2, ..., k. Now, by assumption D3,

$$\hat{\Lambda}(\theta^+) = \Lambda(\theta^*) + o_p(1)$$

where $o_p(1) \to 0$ uniformly in θ^* with probability approaching 1 as $n \to \infty$. Hence, for large enough values of n, $\hat{\Lambda}^{-1}(\theta^+)$ exists and we can write

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = [\Lambda^{-1}(\theta^*) + o_p(1)]\phi \ . \tag{2.2.1}$$

By D6, there exists a nonsingular k x k fixed matrix M such that $M'C(\theta^*)M = I$, the k x k identity matrix. Now consider the i.i.d. random variables

$$Y_i = M'(\frac{\partial}{\partial\theta_1} \log f(X_i, \theta^*), \ \ldots, \ \frac{\partial}{\partial\theta_k} \log f(X_i, \theta^*))' \ ,$$

i = 1, 2, ..., n. These random variables take values in E^k with expectation $E(Y_1) = 0$ and covariance matrix $Cov(Y_1) = I$. Write $M = (m_{ij})_{k \times k}$, i, j = 1, 2, ..., k, $m = \max_{i,j}\{|m_{ij}|\}$ and $L_\ell = \frac{\partial}{\partial\theta_\ell} \log f(X, \theta^*)$, ℓ = 1, 2, ..., k. Let $\rho_4 = E||Y_1||^4$. Then

$$\rho_4 = E[(\textstyle\sum_{j=1}^k m_{1j}L_j)^2 + \ldots + (\sum_{j=1}^k m_{kj}L_j)^2]^2$$
$$\leq m^4k^2 \ E(\textstyle\sum_{j=1}^k L_j)^4.$$

By assumption D5 and Lemma 2.2.1, $\rho_4 < \infty$ and hence, the conditions of Lemma 2.2.3 are satisfied. Let Q_n denote the distribution of $M'\phi$ and let Φ be the standard k-variate normal distribution with covariance matrix I. Then for every real-valued bounded measurable function f on E^k,

$$| \int f \, dQ_n - \int f \, d\Phi |$$
$$\leq \omega_f(E^k) \; a(k) \; \rho_4/\sqrt{n} + (4/3\omega_f^*(2^{7/2}k^{4/3} \; \rho_3\pi^{-1/3}n^{-1/2}: \; \Phi)$$

where a(k) is a constant depending only on k, and $\omega_f(\cdot)$ and $\omega_f^*(\cdot)$ are as defined in Lemma 2.2.3. Since the preceding statement is true for bounded functions, clearly, it is also true for every bounded uniformly continuous function. Since f is bounded, clearly, $\omega_f(E^k)$ is bounded, say, by $K < \infty$. Also, since f is bounded and uniformly continuous, the second term involving ω_f^* can be made as small as we please by choosing n to be large enough independently of θ^*. Thus, given $\varepsilon > 0$, we can choose N (not depending on θ^*) such that

$$| \int f \, dQ_n - \int f \, d\Phi | < \varepsilon \text{ for all } n > N,$$

and for all bounded uniformly continuous functions f on E^k. This shows that

$$Q_n \Rightarrow_u \Phi \, .$$

Let Z denote the standard k-variate normal random variable. Then

$$\phi \Rightarrow_u (M')^{-1} Z$$

and thus,

$$\Lambda^{-1}(\theta^*)\phi \Rightarrow_u \Lambda^{-1}(\theta^*)\ (M')^{-1}Z\ .$$

From this and (2.2.1), it follows that

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \Rightarrow_u N(0,\ \Lambda^{-1}(\theta^*)C(\theta^*)\Lambda^{-1}(\theta^*)\ )$$

where $(M')^{-1}M^{-1} = C(\theta^*)$. //

2.3 Asymptotic Distribution of -2 log λ_n Under Local Alternatives When the Model is Incorrect

Similar to the setting of the estimation problem considered in Section 2.2, we assume in the usual hypothesis testing situation that the model used for constructing test statistics (e.g. the likelihood ratio test statistic -2 log λ_n) has been correctly specified. If this is true, then in many situations where uniformly most powerful test procedures do not exist, the next 'best' procedure to look for may be a locally most powerful test, i.e., we seek a test that performs best against alternatives that are close to the null hypothesis value. In this case, we would then need to derive the relevant asymptotic distributions against the so-called sequence of 'local alternatives'.

To be precise, in ordinary settings where the underlying probability distribution take the form of a family $\{P_\theta, \theta \varepsilon \Theta\}$, it is common to define a sequence of local alternatives as

$$\theta = \theta_0 + \frac{\Delta}{\sqrt{n}}, \; n = 1, 2, \ldots$$

where θ_0 is the hypothesized value of θ and $\Delta = (\Delta_1, \ldots, \Delta_k)'$, is a vector of fixed constants. It is easy to see here that the true value of θ converges to the null hypothesis value θ_0 at a rate of $O(n^{-1/2})$.

Similarly, when the model is misspecified, the same kind of formulation is useful for purposes of hypothesis testing. In this case, the quantity of interest is θ^* as defined in previous sections. It might then be of value to study the performance of the likelihood ratio test, or other tests based on the m.l.e. such as the Rao and the Wald tests, of the hypothesis H_0: $\theta^* = \theta_0$ against a sequence of local alternatives of the form

$$\theta^* = \theta_0 + \frac{\Delta}{\sqrt{n}}, \quad n = 1, 2, \ldots$$

where as before, $\Delta = (\Delta_1, \ldots, \Delta_k)'$ is a vector of fixed constants. In order to carry out such an investigation, it is then necessary to derive the asymptotic distribution of any such test statistic under the afore-mentioned local alternatives when the model is incorrect.

To establish the above asymptotic distribution, we need the following lemma.

Lemma 2.3.1. Let T, a p x 1 random vector, have the p-variate normal distribution with mean vector μ and a positive definite covariance matrix $\sum$. Define the quadratic form

$$Q = T'AT,$$

where A is a p x p matrix of fixed constants. Then Q is distributed as a linear combination of i.i.d. noncentral chi-squared random variables:

$$Q = \sum_{i=1}^{p} \lambda_i \chi_i^2 (1, \mu' \Sigma^{-1}\mu)$$

where $\chi_i^2(1, \mu' \Sigma^{-1}\mu)$, $i = 1, 2, \ldots, p$ are independent and identically distributed noncentral chi-squared random variables with 1 d.f. and noncentrality parameter $\mu' \Sigma^{-1} \mu$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of the matrix ΣA.

<u>Proof</u>. Without loss of generality, we can assume that A is symmetric. Since Σ is positive definite, there exists a nonsingular matrix L such that $\Sigma = LL'$. Let $Y = L^{-1}T$. Then $Y \sim N(L^{-1} \mu, I)$ and $Q = Y'L'ALY$. Since $L'AL$ is symmetric, there exists an orthogonal matrix C such that $C'L'ALC$ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ of $L'AL$. Let $Z = C'Y$. Then

$$Z \sim N(C'L^{-1}\mu, I) \quad \text{and}$$

$$\begin{aligned} Q &= (CZ)' L'AL(CZ) \\ &= Z' C' L' AL CZ \\ &\overset{d}{\sim} \sum_{i=1}^{p} \lambda_i \chi_i^2 (1, \delta) \end{aligned}$$

where $\chi_i^2(1, \delta)$, $i = 1, 2, \ldots, p$ are i.i.d. noncentral chi-squared random variables with 1 d.f. and noncentrality parameter $\delta = (C' L^{-1}\mu)' (C'L^{-1}\mu) = \mu' \Sigma^{-1} \mu$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eignevalues of the matrix $L'AL$. To complete the proof, we have to show that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are also the eigenvalues of the matrix ΣA. However, this follows easily since

$$|L'AL - \lambda I| = 0$$

$$\iff |L| \cdot | L'AL - \lambda I| \cdot |L^{-1}| = 0$$

$$\iff |LL' ALL^{-1} - \lambda I| = 0$$

$$\iff |\Sigma A - \lambda I| = 0 \ . \quad //$$

Before we proceed to state and prove the next theorem, we will need to modify the earlier regularity conditions C1 - C4 of Chapter 1. All convergences in probability and in distribution implied in C1 - C4 are now assumed to hold uniformly in $\theta^* \varepsilon \Theta$. We shall, for clarity, rename this new set of regularity conditions as C1' - C4'.

<u>Theorem 2.3.1</u>. Assume the regularity conditions C1' - C4' hold. Under the local alternatives

$$\theta^* = \theta_0 + \frac{\Delta}{\sqrt{n}} \quad , \ n = 1, 2, \ldots$$

where $\Delta = (\Delta_1, \ldots, \Delta_k)'$, is a vector of constants, the likelihood ratio statistic, T_n, is asymptotically distributed as a linear combination of noncentral chi-squared random variables:

$$T_n = - 2 \log \lambda_n \xrightarrow{D} \sum_{i=1}^{k} C_i' \chi_i^2(1, \beta)$$

where $\chi_i^2(1, \beta)$, i = 1, 2, ..., k are i.i.d. noncentral chi-squared random variables with 1 degree of freedom and noncentrality parameter $\beta = (\Lambda\Delta)' C^{-1}(\theta_0)(\Lambda\Delta)$ and $C_1' \geq C_2' \geq \ldots \geq C_k'$ are the eignevalues of the matrix $-\Lambda^{-1}C(\theta_0)$.

<u>Proof</u>. We write

$$- 2 \log \lambda_n = 2[\sum_{i=1}^{n} \log f(X_i, \hat{\theta}_n) - \sum_{i=1}^{n} \log f(X_i, \theta_0)]$$

and expand each term in the parenthesis about θ^*. (This is possible because of C_2'.)

$$\sum_{i=1}^{n} \log f(X_i, \hat{\theta}_n)$$
$$= \sum_{i=1}^{n} \log f(X_i, \theta^*) + \sum_{\ell=1}^{k} (\hat{\theta}_{n\ell} - \theta_\ell^*) \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} f(X_i, \theta^*)$$
$$+ \frac{1}{2} \sum_{\ell,m=1}^{k} (\hat{\theta}_{n\ell} - \theta_\ell^*)(\hat{\theta}_{nm} - \theta_m^*) \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta^+)$$

where $\theta^+ = \theta^* + \alpha(\hat{\theta}_n - \theta^*)$, $0 \leq \alpha \leq 1$ and

$$\sum_{i=1}^{n} \log f(X_i, \theta_0)$$
$$= \sum_{i=1}^{n} \log f(X_i, \theta^*) + \sum_{\ell=1}^{k} (\theta_{0\ell} - \theta_\ell^*) \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta^*)$$
$$+ \frac{1}{2} \sum_{\ell,m=1}^{k} (\theta_{0\ell} - \theta_\ell^*)(\theta_{0m} - \theta_m^*) \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta^{++})$$

where $\theta^{++} = \theta^* + \beta(\theta_0 - \theta^*)$, $0 \leq \beta \leq 1$.

So $- 2 \log \lambda_n$

$$= 2\sum_{\ell=1}^{k} (\hat{\theta}_{n\ell} - \theta_{0\ell}) \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta^*)$$
$$+ \sum_{\ell,m=1}^{k} (\hat{\theta}_{n\ell} - \theta_\ell^*)(\hat{\theta}_{nm} - \theta_m^*) \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta^+)$$
$$+ \sum_{\ell,m=1}^{k} (\theta_{0\ell} - \theta_\ell^*)(\theta_{0m} - \theta_m^*) \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell\partial\theta_m} \log f(X_i, \theta^{++}) .$$

By assumption C2', $\hat{\theta}_n \xrightarrow{p} \theta^*$ uniformly in θ^*. Also under

The local alternatives $\theta^* = \theta_0 + \Delta/\sqrt{n}$, $\theta^* \to \theta_0$ as $n \to \infty$. This implies that $\theta^+ \xrightarrow{p} \theta_0$ and $\theta^{++} \xrightarrow{p} \theta_0$. Assumption C2' also implies that $\sqrt{n}(\hat{\theta}_{n\ell} - \theta_\ell^*)$ converges in distribution uniformly in θ^* for each $\ell = 1, 2, \ldots, k$ and assumption C4' implies that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X_i, \theta^*) \xrightarrow{p} E[\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X, \theta^*)] = \Lambda_{\ell m}$$

uniformly in θ^* for every ℓ, $m = 1, 2, \ldots, k$. By Slutsky's Theorem, the previous expression can be rewritten as

$$- 2 \log \lambda_n \overset{\text{a.d.}}{=\!=} 2 \sum_{\ell=1}^{k} \{\sqrt{n}(\hat{\theta}_{n\ell} - \theta_\ell^*) + \Delta_\ell\} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta^*)$$

$$+ \sum_{\ell,m=1}^{k} \{ \sqrt{n}(\hat{\theta}_{n\ell} - \theta_\ell^*) \sqrt{n}(\hat{\theta}_{nm} - \theta_m^*0 - \Delta_\ell \Delta_m\} E[\frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X, \theta^*)]$$

A similar expansion yields

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i \ \theta^*)$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \hat{\theta}_n) + \sum_{\ell=1}^{k} (\theta_\ell^* - \hat{\theta}_{n\ell}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X_i \bar{\theta})$$

where $\bar{\theta} = \hat{\theta}_n + \gamma(\theta^* - \hat{\theta}_n)$, $0 \leq \gamma \leq 1$. The first term vanishes by the maximizing property of m.l.e. By assumption C2' $\hat{\theta}_n \xrightarrow{p} \theta^*$ and so $\bar{\theta} \xrightarrow{p} \theta^*$. Hence

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta^*)$$

$$\overset{\text{a.d.}}{=\!=} -\sum_{\ell=1}^{k} \sqrt{n}(\hat{\theta}_{n\ell} - \theta_\ell^*) \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta_\ell \partial\theta_m} \log f(X_i, \theta^*)$$

$$\overset{\text{a.d.}}{=\!=} -\sum_{\ell=1}^{k} \sqrt{n}(\hat{\theta}_{n\ell} - \theta_\ell^*) \Lambda_{\ell m} .$$

Let $Z = \sqrt{n}(\hat{\theta}_n - \theta^*)$. Then

$$- 2 \log \lambda_n \xrightarrow{\text{a.d.}} \{2\ (Z + \Delta)'\ (-\Lambda)Z\}$$

$$+ (Z + \Delta)'\Lambda\ (Z - \Delta)$$

$$= (Z + \Delta)'\ (-\Lambda)\ (Z + \Delta).$$

Since $Z + \Delta \xrightarrow{D} N(\Delta, \Lambda^{-1}C(\theta_0)(\Lambda')^{-1})$ by assumption C2, an application of Lemma 2.3.1 yields the desired result:

$$- 2 \log \lambda_n \xrightarrow{D} \sum_{i=1}^{k} C_i'\chi^2_i\ (1, \delta)$$

where $\chi_i^2\ (1, \delta)$, $i = 1, 2, \ldots, k$ are i.i.d. chi-squared random variables with 1 d.f. and noncentrality parameter

$$\delta = \Delta'(\Lambda^{-1}C(\theta_0)(\Lambda')^{-1})^{-1}\Delta$$

$$= (\Lambda\Delta)'\ C^{-1}(\theta_0)(\Lambda\Delta)\ ,$$

and $C_1' \geq C_2' \geq \ldots \geq C_k'$ are the eigenvalues of the matrix $\Lambda^{-1}C(\theta_0)(\Lambda')^{-1}(-\Lambda) = -\Lambda^{-1}C(\theta_0)$. //

<u>Remark</u>. When the model is correct, $-\Lambda = C(\theta_0)$. So $C_1' = C_2' = \ldots = C_k' = 1$, $\delta = \Delta'\ C(\theta_0)\Delta$ and $-2 \log \lambda_n$ is asymptotically distributed (under the sequence of local alternatives) as a noncentral chi-squared raondm variable with k degrees of freedom and noncentrality parameter $\Delta'\ C(\theta_0)\Delta$ which agrees with known results.

CHAPTER 3

ASYMPTOTIC DISTRIBUTION OF THE RAO AND WALD STATISTICS UNDER MODEL MISSPECIFICATION

In Chapter 1, we discussed three important hypothesis testing procedures, namely, the tests based on the likelihood ratio test statistic, the Rao statistic and the Wald statistic. Foutz and Srivastava (1978) derived the asymptotic distribution of the likelihood ratio test statistic when the model is incorrect.

In this chapter, we shall derive the corresponding asymptotic distributions of the Rao statistic and the Wald statistic under model misspecification.

3.1 Asymptotic Distribution of the Rao Statistic Under Model Misspecification

Before we proceed with the asymptotics, we will recall again the expression for the Rao statistic R_n. Let $\theta' = (\delta', \gamma')$, $\tilde{\theta}_n' = (\tilde{\delta}_n', \tilde{\gamma}_n')$, $\delta' = (\theta_1, \ldots, \theta_r)$, $\gamma' = (\theta_{r+1}, \ldots, \theta_k)$, $\tilde{\delta}_n' = (\theta_1, \ldots, \theta_r)$, $\tilde{\gamma}_n' = (\theta_{r+1}, \ldots, \theta_k)$, $1 \leq r \leq k$ and

$$I(\theta) = \left(E_\theta[\frac{\partial}{\partial\theta_\ell} \log f(X, \theta) \cdot \frac{\partial}{\partial\theta_m} \log f(X, \theta)] \right)_{k \times k}$$

$\ell, m = 1, 2, \ldots, k$. Define

$$\begin{aligned} \Theta_0 &= \{\theta |\ \theta\varepsilon\Theta,\ \theta_j = \theta_{oj},\ 1 \leq j \leq r\} \\ &= \{\theta |\ \theta\varepsilon\Theta,\ \delta = \delta_0\}\ , \end{aligned}$$

where $\delta_0 = (\theta_{01}, \ldots, \theta_{or})'$ is a r x 1 vector of fixed constants. Let $\tilde{\theta}_n$ denote the restricted m.l.e. of θ over Θ_0 and let

$$V_\theta = (\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_1} \log f(X_i, \theta), \ldots, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_k} \log f(X_i, \theta))' .$$

For testing the hypothesis

$$H_0:\ \theta \varepsilon \Theta_0 \quad \text{versus} \quad H_1:\ \theta \varepsilon \Theta - \Theta_0,$$

the Rao statistic is defined as

$$R_n = V_{\tilde{\theta}_n}{}' \, I^{-1}(\tilde{\theta}_n) V_{\tilde{\theta}_n} \quad .$$

To avoid unnecessary complications and to afford easy comparisons, we shall use the same notations and assume the same regularity conditions as in Foutz and Srivastava (1978), namely C1 - C4, as well as the following condition C5, to prove the main theorems in this chapter.

C5. Assume that the third order derivatives $\frac{\partial^3}{\partial\theta_\ell \partial\theta_m \partial\theta_n} \log f(x, \theta)$, $\ell, m, n = 1, 2, \ldots, k$ exist and are finite in a neighborhood of θ^*.

Theorem 3.1.1. Under the regularity conditions C1 - C5, and if $\theta^* \in \Theta_0$, then R_n is asymptotically distributed as a linear combination of k i.i.d. chi-squared random variables:

$$R_n \xrightarrow{D} \sum_{i=1}^{k} \ell_i \chi_i^2$$

where $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_k$ are the eigenvalues of

$$C(\theta^*)[I^{-1}(\theta^*) - \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} (\Lambda_{22}')^{-1} L \Lambda_{22}^{-1}(0' \; I_{k-r})]$$

and $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$ are k i.i.d. chi-squared random variables with 1 degree of freedom. (Here I_{k-r} is the (k-r)x(k-r) identity matrix and 0 is the rx (k-r) matrix of zeros.)

<u>Remark</u>. It is interesting to note that if the hypothesis is simple, i.e., $H_0 = \theta = \theta_0$, then the proof of this theorem is straight forward. In this case, $\theta^* = \theta_0$. So $V_{\theta_0} \xrightarrow{D} N(0, C(\theta_0))$ by the Multivariate Central Limit Theorem and by Lemma 2.3.1,

$$R_n = V_{\theta_0}' I^{-1}(\theta_0)V_{\theta_0} \xrightarrow{D} \sum_{i=1}^{k} \ell_i \chi_i^2 ,$$

where $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$ are k i.i.d. chi-squared random variables with 1 degree of freedom and $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_k$ are the eigenvalues of $C(\theta_0)I^{-1}(\theta_0)$. In particular, if the model has been correctly specified, then $C(\theta_0) = I(\theta_0)$, implying that $\ell_1 = \ell_2 = \ldots \ell_k = 1$ and thus R_n is asymptotically distributed as a chi-squared random variable with k degrees of freedom.

<u>Proof of Theorem 3.1.1</u>. Write $\tilde{\gamma}_n = (\tilde{\gamma}_{n1}, \ldots, \tilde{\gamma}_{n(k-r)})'$. under the assumption $\theta^* \varepsilon \Theta_0$, $\theta^* = (\delta_0', \gamma^{*\prime})'$ and we can regard $V_{\theta^*}' I^{-1}(\theta^*)V_{\theta^*}$ as a function of $\gamma^* = (\theta^*_{r+1}, \ldots, \theta^*_k)'$ only. Expanding this function about $\tilde{\theta}_n$ (i.e. $\tilde{\gamma}_n$),

$$\begin{aligned} &V_{\theta^*}' I^{-1}(\theta^*) V_{\theta^*} \\ &= R_n + \sum_{j=1}^{k-r} (\gamma_j^* - \tilde{\gamma}_{nj}) \frac{\partial}{\partial\gamma_j} (V_\theta' I^{-1}(\theta)V_\theta)\Big|_{\theta=\tilde{\theta}_n} \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{k-r} (\gamma_i^* - \tilde{\gamma}_{ni})(\gamma_j^* - \tilde{\gamma}_{nj}) \frac{\partial^2}{\partial\gamma_i\partial\gamma_j} (V_\theta' I^{-1}(\theta)V_\theta)\Big|_{\theta=\theta^+} \end{aligned}$$

where $\theta^+ = \tilde{\theta}_n + \lambda(\theta^* - \tilde{\theta}_n)$, $0 \leq \lambda \leq 1$. First, let us consider the second term. Write

$\phi_\ell = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i, \theta)$, $\ell = 1, 2, \ldots, k$ and

$I^{-1}(\theta) = (I^{\ell m})_{k \times k}$, $\ell, m = 1, 2, \ldots, k$. Hence, $V_\theta = (\phi_1, \ldots, \phi_k)'$ and

$$V_\theta' \; I^{-1}(\theta) \; V_\theta = \sum_{\ell,m=1}^{k} I^{\ell m} \; \phi_\ell \phi_m \; ,$$

$$\frac{\partial}{\partial\gamma_j} (V_\theta' \; I^{-1}(\theta) V_\theta) = \sum_{\ell,m=1}^{k} \frac{\partial}{\partial\gamma_j} (I^{\ell m} \; \phi_\ell \phi_m) \; .$$

where $\frac{\partial}{\partial\gamma_j} (I^{\ell m} \; \phi_\ell \phi_m)$

$$= (\frac{\partial I^{\ell m}}{\partial\gamma_j}) \; \phi_\ell \phi_m + I^{\ell m} \; (\frac{\partial\phi_\ell}{\partial\gamma_j}) \; \phi_m + I^{\ell m} \; \phi_\ell (\frac{\partial\phi_m}{\partial\gamma_j}) \; .$$

By the definition of restricted m.l.e. $\tilde{\theta}_n$, under H_0,

$$\phi_\ell \Big|_{\theta=\tilde{\theta}_n} = 0, \; \ell = 1, 2, \ldots, k \; .$$

So $\frac{\partial}{\partial\gamma_j} (I^{\ell m}\phi_\ell\phi_m) \Big|_{\theta=\tilde{\theta}_n} = 0$ for each $j = 1, 2, \ldots, k-r$ and

$$\frac{\partial}{\partial\gamma_j} V_\theta' \; I^{-1}(\theta) V_\theta \Big|_{\theta=\tilde{\theta}_n} = \sum_{\ell,m=1}^{k} \frac{\partial}{\partial\gamma_j} (I^{\ell m}\phi_\ell\phi_m) \Big|_{\theta=\tilde{\theta}_n} = 0 \; .$$

Next, we consider the third term. By direct computation,

$$\frac{\partial^2}{\partial\gamma_i\partial\gamma_j}(V_\theta'\ I^{-1}(\theta)\ V_\theta)\Big|_{\theta=\theta^+}$$

$$= \sum_{\ell,m=1}^{k} \{[(\frac{\partial^2 I^{\ell m}}{\partial\gamma_i\partial\gamma_j}\ \phi_\ell\phi_m + (\frac{\partial I^{\ell m}}{\partial\gamma_j})(\frac{\partial\phi_\ell}{\partial\gamma_i})\ \phi_m$$

$$+ (\frac{\partial I^{\ell m}}{\partial\gamma_j})\ \phi_\ell\ (\frac{\partial\phi m}{\partial\gamma_i})] + [(\frac{\partial I^{\ell m}}{\partial\gamma_i})(\frac{\partial\theta_\ell}{\partial\gamma_j})\ \phi_m$$

$$+ I^{\ell m}(\frac{\partial^2\phi_\ell}{\partial\gamma_i\partial\gamma_j})\ \phi_m + I^{\ell m}(\frac{\partial\phi_\ell}{\partial\gamma_j})(\frac{\partial\phi_m}{\partial\gamma_i})]$$

$$+ [(\frac{\partial I^{\ell m}}{\partial\gamma_i})\ \phi_\ell\ (\frac{\partial\phi_m}{\partial\gamma_j}) + I^{\ell m}(\frac{\partial\phi_\ell}{\partial\gamma_i})(\frac{\partial\phi_m}{\partial\gamma_j})$$

$$+ I^{\ell m}\ \phi_\ell(\frac{\partial^2\phi_m}{\partial\gamma_i\partial\gamma_j})]\}\Big|_{\theta=\theta^+}$$

Now $(\gamma_i^* - \tilde{\gamma}_{ni})(\gamma_j^* - \tilde{\gamma}_{nj})[(\frac{\partial^2 I^{\ell m}}{\partial\gamma_i\partial\gamma_j})\phi_\ell\phi_m]\Big|_{\theta=\theta^+}$

$$= \sqrt{n}\ (\gamma_i^* - \tilde{\gamma}_{ni})\sqrt{n}(\gamma_j^* - \tilde{\gamma}_{nj})[(\frac{\partial^2 I^{\ell m}}{\partial\gamma_i\partial\gamma_j})\ \frac{\phi_\ell}{\sqrt{n}}\ \frac{\phi_m}{\sqrt{n}}\ ]\Big|_{\theta=\theta^+}$$

$$\xrightarrow{\ p\ } 0 \text{ since } \frac{\phi_\ell}{\sqrt{n}}\Big|_{\theta=\theta^+} \xrightarrow{\ p\ } 0$$

(because $\theta^+ \xrightarrow{\ p\ } \theta^*$ and as a consequence of the Weak Law of Large Numbers) and $\sqrt{n}(\gamma_i^* - \tilde{\gamma}_{ni})$ converges in distribution by assumption C2. In a similar fashion, all other terms converge to zero in probability except terms of the form

$$I^{\ell m}[\frac{\partial\phi_\ell}{\partial\gamma_i} \frac{\partial\phi_m}{\partial\gamma_j} + \frac{\partial\phi_\ell}{\partial\gamma_j} \frac{\partial\phi_m}{\partial\gamma_i}] .$$

So, with probability 1 as $n \to \infty$, we reduce the third term to

$$\frac{1}{2} \sum_{i,j=1}^{k-r} (\gamma_i^* - \tilde{\gamma}_{ni})(\gamma_j^* - \tilde{\gamma}_{n_j}) \sum_{\ell,m=1}^{k} I^{\ell m}[\frac{\partial\phi_\ell}{\partial\gamma_j} \frac{\partial\phi_m}{\partial\gamma_i} + \frac{\partial\phi_\ell}{\partial\gamma_i} \frac{\partial\phi_m}{\partial\gamma_j}]\Big|_{\theta=\theta^+}$$

Let $W(i,j) = \frac{\partial\phi_i}{\partial\theta_j}$, $i, j = 1, 2, \ldots, k$ and $W = (W(i,j))_{k x k}$. The preceeding expression then becomes

$$\frac{1}{2} \sum_{i,j=1}^{k-r} (\gamma_k^* - \tilde{\gamma}_{ni})(\gamma_j^* - \tilde{\gamma}_{nj}) \sum_{\ell,m=1}^{k} I^{\ell m} W(\ell,j+r)W(m,i+r)$$

$$+ \sum_{\ell,m=1}^{k} I^{\ell m}W(\ell, i+r)W(m,j+r)\}\Big|_{\theta=\theta^+}$$

$$= \frac{1}{2} \sum_{i,j=1}^{k-r} (\gamma_i^* - \tilde{\gamma}_{ni})(\gamma_j^* - \tilde{\gamma}_{nj})\{\underset{\sim}{W}_{j+r}' I^{-1}(\theta) \underset{\sim}{W}_{i+r}$$

$$+ \underset{\sim}{W}_{i+r}' I^{-1}(\theta)\underset{\sim}{W}_{j+r}\}\Big|_{\theta=\theta^+}$$

where $W = (\underset{\sim}{W}_1, \ldots, \underset{\sim}{W}_k)$ and $\underset{\sim}{W}_p$, $p = 1, 2, \ldots, k$ are $k \times 1$ vectors. Note that $\underset{\sim}{W}_i' I^{-1}(\theta) \underset{\sim}{W}_j$ is the (i,j)th element of $W'I^{-1}(\theta)W$ and by symmetry,

$$\underset{\sim}{W}_j' I^{-1}(\theta)\underset{\sim}{W}_i = \underset{\sim}{W}_i' I^{-1}(\theta)\underset{\sim}{W}_j .$$

Let $B(\theta)$ be the lower $(k-r) \times (k-r)$ diagonal block of $W'I^{-1}(\theta)W$. The last expression can then be compactly written as

$$\sqrt{n}\,(\tilde{\gamma}_n - \gamma^*)' \left[\frac{B(\theta+)}{n}\right] \sqrt{n}\,(\tilde{\gamma}_n - \gamma^*).$$

By assumption C2, $\tilde{\theta}_n \overset{P}{\to} \theta^*$. This implies $\theta^+ \overset{P}{\to} \theta^*$. Also $\sqrt{n}(\tilde{\gamma}_n - \gamma^*)$ converges in distribution. Further, assumption C2 and C4 imply

$$\frac{B(\theta+)}{n} \overset{P}{\to} L$$

where L is the lower $(k-r)x(k-r)$ diagonal block of $\Lambda' I^{-1}(\theta^*)\Lambda$. Thus, by applying Slutsky's Theorem, we can write

$$R_n \overset{a.d.}{=} V_{\theta^*}' I^{-1}(\theta^*) V_{\theta^*} - n(\tilde{\gamma}_n - \gamma^*)' L (\tilde{\gamma}_n - \gamma^*)$$

where "$\overset{a.d.}{=}$" denotes "equal in asymptotic distribution." By a similar expansion, letting $U_\gamma = (\phi_{r+1}, \ldots, \phi_k) = (0\ I_{k-r})V_\theta$ where 0 represents the $(k-r)xr$ matrix of zeros and I_{k-r} is the $(k-r) \times (k-r)$ identity matrix, we have

$$U_{\gamma^*} = U_{\tilde{\gamma}_n} + \left| \begin{matrix} \Sigma_{j=1}^{k-r} (\gamma_{j^*} - \tilde{\gamma}_{nj}) \frac{\partial \phi r+1}{\partial \gamma_j} \\ \vdots \\ \Sigma_{j=1}^{k-r} (\gamma_{j^*} - \gamma_{nj}) \frac{\partial \phi k}{\partial \gamma_j} \end{matrix} \right|_{\theta = \theta^{++}}$$

where $\theta^{++} = \tilde{\theta}_n + \beta(\theta^* - \tilde{\theta}_n)$, $0 \leq \beta \leq 1$. Alternatively, since $U_{\tilde{\gamma}n} = 0$ by the definition of the m.l.e. $\tilde{\theta}_n$, under Ho,

$$U_{\gamma*} \overset{a.d.}{=} \Lambda_{22} \sqrt{n}\,(\gamma^* - \tilde{\gamma}_n)$$

(since $\tilde{\theta}_n \overset{P}{\to} \theta^*$ implies $\theta^{++} \overset{P}{\to} \theta^*$). So we have

$$\sqrt{n}\,(\gamma^* - \tilde{\gamma}_n) \overset{a.d.}{=} \Lambda_{22}^{-1} U_{\gamma*} = \Lambda_{22}^{-1}\,(0\ I)\ V_{\theta*}.$$

Hence,

$$R_n \overset{a.d.}{=} V_{\theta*} \left[I^{-1}(\theta^*) - \begin{pmatrix} 0' \\ I_{k-r} \end{pmatrix} (\Lambda_{22}')^{-1} L\, \Lambda_{22}^{-1} (0'\ \ I_{k-r}) \right] V_{\theta*}$$

By the Multivariate Central Limit Theorem,

$$V_{\theta*} \overset{D}{\to} N(0,\ C(\theta^*)\ ).$$

Applying Lemma 2.3.1, it follows that

$$R_n \overset{a.d.}{=} \Sigma_{i=1}^{k}\ \ell_i \chi_i^2$$

where $\chi_1^2, \ldots, \chi_k^2$ are i.i.d. chi-squared random variables with 1 d.f. and $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_k$ are the eigenvalues of

$$C(\theta^*) \left[I^{-1}(\theta^*) - \begin{pmatrix} 0' \\ I_{k-r} \end{pmatrix} (\Lambda_{22}')^{-1} L\, \Lambda_{22}^{-1} (0'\ \ I_{k-r}) \right].$$

<u>Remark</u>. When the model is correct, we have

$C(\theta^*) = I(\theta^*) = -\Lambda$, $L = -\Lambda_{22}$ and so the above matrix expression reduces to

$$I - \begin{pmatrix} 0 & \Lambda_{12}'\ \Lambda_{22} \\ 0 & I_{k-r} \end{pmatrix} = \begin{pmatrix} I_r & -\Lambda_{12}'\ \Lambda_{22} \\ 0 & 0 \end{pmatrix}$$

It is clear that $\ell_1 = .. = \ell_r = 1$ and $\ell_{r+1} = \ldots = \ell_k = 0$ and so R_n has an asymptotic chi-square distribution with r d.f. When the model is incorrect, it is our conjecture that there are only r nonzero roots. However, we are unable to establish this at this point.

We indicate the "non-null" behavior of R_n in the following theorem.

Theorem 3.1.2: Assume the existence of a unique θ_o^* that maximizes $E[\log f(X,\theta)]$ over Θ_o, and assume that the regularity conditions C1-C2 hold. Also assume that for some open neighborhood Θ_o^* about θ_o^*,

$$\sup_{\theta\in\Theta_o^*} \left| \frac{1}{n}\Sigma_{i=1}^n \frac{\partial}{\partial\theta_\ell} \log f(X_i,\theta) - E[\frac{\partial}{\partial\theta_\ell} \log f(X,\theta)] \right| \overset{P}{\to} 0 ,$$

$\ell = 1, 2, \ldots, k$. If $\theta^* \notin \Theta_o$, then

$$\frac{R_n}{n} \overset{P}{\to} E(V_{\theta_o^*})' \, I^{-1}(\theta_o^*) \, E(V_{\theta^*})$$

where $E(V_{\theta_o^*}) = (E[\frac{\partial}{\partial\theta_1}\log f[X,\theta_o^*], \ldots , E[\frac{\partial}{\partial\theta_k}\log f(X,\theta_o^*)])'$.

Proof: Let $\ell_n(\theta) = (\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta_1} \log f(X_i,\theta), \ldots , \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta_k}\log f(X_i,\theta))'$. Then since $\tilde{\theta}_n \overset{P}{\to} \theta_o^*$, $\ell_n(\tilde{\theta}_n) \overset{P}{\to} EV_{\theta_o^*}$ by the Weak Law of Large Numbers. Also $I^{-1}(\tilde{\theta}_n) \overset{P}{\to} I^{-1}(\theta_o^*)$ and the result follows.

Remark: Since $I^{-1}(\theta_o^*)$ is positive definite, this establishes the consistency of the Rao test which rejects for large values of R_n (because $\frac{R_n}{n}$ converge in probability to a positive constant under $\theta^* \notin \Theta_o$).

3.2 Asymptotic Distribution of the Wald Statistic Wn Under Model Misspecification

Partition the unrestricted m.l.e. $\hat{\theta}_n' = (\hat{\delta}_n' \ \ \hat{\gamma}_n')$. The Wald statistic W_n is defined as

$$W_n = n(\hat{\delta}_n - \delta_o)' \left[(I_r \ 0)\, I^{-1}(\hat{\theta}_n) \binom{I_r}{0'}\right]^{-1}(\hat{\delta}_n - \delta_o)$$

where I_r is the r x r identity matrix and 0 is the r x (k-r) matrix of zeros. Let Θ_o be as defined in the previous section.

Theorem 3.2.1: Under the regularity conditions C1-C2 and if $\theta^* \varepsilon\ \Theta_o$, then W_n is asymptotically distributed as a linear combination of r i.i.d. chi-squared random variables:

$$W_n \xrightarrow{D} \sum_{i=1}^{r} d_i\, \chi_i^2$$

where $d_1 \geq d_2 \geq \ldots \geq d_r$ are the eigenvalues of $M[(I_r \ 0)\, I^{-1}(\theta^*) \binom{I_r}{0'}]^{-1}$, M being the upper r x r diagonal block of $\Lambda^{-1}C(\theta^*)(\Lambda')^{-1}$, and $\chi_1^2, \ldots, \chi_r^2$ are r i.i.d. chi-squared random variables with 1 degree of freedom (df).

Proof: By assumption C2, $\hat{\theta}_n \xrightarrow{P} \theta^*$. Thus $I^{-1}(\hat{\theta}_n) \xrightarrow{P} I^{-1}(\theta^*)$. Also $\sqrt{n}\,(\hat{\delta}_n - \delta_o) \xrightarrow{D} N(0, M)$ where M is the upper diagonal block of $\Lambda^{-1}C(\theta^*)(\Lambda')^{-1}$. By Lemma 2.3.1 and also by the application of Slutky's Theorem,

$$W_n \overset{a.d.}{=} n(\hat{\delta}_n - \delta_o)' \left[(I_r \ 0)\, I^{-1}(\theta^*) \binom{I_r}{0'}\right]^{-1}(\hat{\delta}_n - \delta_o) \overset{a.d.}{=} \sum_{i=1}^{r} d_i \chi_i^2$$

where $\chi_1^2, \ldots, \chi_r^2$ are independent and identically distributed chi-squared random variables with 1 df and $d_1 \geq d_2 \geq \ldots \geq d_r$ are the eigenvalues of

$$M \, [(I_r \;\; 0) \; I^{-1} \, (\theta^*) \, \binom{I_r}{0'} \,]^{-1}.$$

The non-null behavior of W_n is indicated in

<u>Theorem 3.2.2</u>: Under assumptions C1-C2 and if $\theta^* \notin \Theta_o$, then

$$\frac{W_n}{n} \xrightarrow{P} (\delta^* - \delta_o)' \, [(I_r \;\; 0) \; I^{-1}(\theta^*) \, \binom{I_r}{0'} \,]^{-1}(\delta^* - \delta_o).$$

<u>Proof</u>: The proof follows immediately from assumption C2.

<u>Remark</u>: This establishes the 'consistency' of the Wald test for the same reasons as in the case of the Rao test.

3.3 Examples

We will conclude this chapter by considering two examples.

<u>Example 3.3.1</u>: Consider the family of Poisson distributions with mean $\theta>0$ as the model chosen for the purpose of testing the simple hypothesis that the mean of the distribution of X is a specified constant θ_o. The m.l.e. for θ is $\overline{X}$ and the likelihood ratio statistic constructed from this model is $-2 \log \lambda_n$ where

$$\lambda_n = (\theta_o/\overline{X})^{n\overline{X}} \; \exp \{n(\overline{X} - \theta_o)\} \; .$$

Direct computation yields $I(\theta) = 1/\theta$ and so the Rao and the Wald statistics take the following forms:

$$R_n = n(\overline{X} - \theta_o)^2/\theta_o$$

$$W_n = n(\overline{X} - \theta_o)^2/\overline{X}.$$

[Note that in this example, $W_n = \theta_o R_n/(\theta_o R_n/n + \theta_o)^{1/2}$.]
When the Poisson model is incorrect and X has a discrete distribution $\{P_n\}_{n=0}^{\infty}$ satisfying assumptions C1-C5 (e.g. the binomial distribution with parameters k (a positive integer) and p $(0 < p < 1)$, then it can be easily shown that

$\theta^* = \theta_o$, $= -W = -1/\theta_o$, $C(\theta^*) = \mathrm{Var}(X)/\theta_o^2$,

$M = \mathrm{Var}(X)$ where $\mathrm{Var}(X)$ denotes the variance of X, so $-2\log\lambda_n$ is asymptotically distributed as $[\mathrm{Var}(X)/\theta_o]\,\chi_1^2$.

Similarly, $I(\theta^*) = 1/\theta_o$ and so W_n has the same asymptotic distribution $[\mathrm{Var}(X)/\theta_o]\cdot\chi_1^2$. Also $C(\theta^*)\,I^{-1}(\theta^*) = [\mathrm{Var}(X)/\theta_o^2]\,[1/\theta_o]^{-1}$ $= [\mathrm{Var}(X)/\theta_o]$ and the same asymptotic distribution holds true for R_n. However, if $\theta^* \neq \theta_o$, all three statistics have different probability limits, viz.,

$$-\frac{2}{n}\log\lambda n \xrightarrow{P} 2[(\theta_o - \theta^*) + \theta^*(\log\theta^* - \log\theta_o)],$$

$$\frac{Rn}{n} \xrightarrow{P} (\theta^* - \theta_o)^2/\theta_o,$$

$$\frac{Wn}{n} \xrightarrow{P} (\theta^* - \theta_o)^2/\theta^*.$$

Example 3.3.2: Consider the problem of testing that a random variable X has a specified variance σ_o^2. Suppose the family of normal distributions with variance $\theta_1 > 0$ and unknown finite mean θ_2 is chosen as the model for constructing the tests. The restricted m.l.e. of $\theta = (\theta_1, \theta_2)'$, $\tilde{\theta}_n$, is $(\sigma_o^2\ \bar{X})'$, and the unrestricted m.l.e. $\tilde{\theta}_n$ is $(S_n\ \bar{X})'$ where S_n is the sample variance. Suppose the normal model is incorrect and instead, X has a uniform distribution on the interval (α, β), then assumptions C1-C4 are satisfied with

$\theta^* = ([\beta - \alpha]^2/12,\ [\beta + \alpha]/2)'$.

Direct computation yields

$$N = \begin{bmatrix} -72/(\beta - \alpha)^4 & 0 \\ 0 & -12/(\beta - \alpha)^2 \end{bmatrix}$$

$$C(\theta^*) = \begin{bmatrix} 28.8/(\beta - \alpha)^4 & 0 \\ 0 & 12/(\beta - \alpha)^2 \end{bmatrix}$$

and hence, $\Lambda^{-1}C(\theta^*)(\Lambda')^{-1} = \begin{bmatrix} (\beta - \alpha)^4/180 & 0 \\ 0 & (\beta - \alpha)^2/12 \end{bmatrix}$

Thus $M = (\beta - \alpha)^4/180$, $W = 72/(\beta - \alpha)^4$ and $-2\log\lambda n$ is asymptotically distributed as $.4\chi_1^2$.

To obtain the asymptotic distributions of R_n and W_n, we first compute

$$I(\theta^*) = \begin{bmatrix} 72/(\beta - \alpha)^4 & 0 \\ 0 & 12/(\beta - \alpha)^2 \end{bmatrix}$$

So $\Lambda' I^{-1}(\theta^*)\Lambda = \begin{bmatrix} 72/(\beta - \alpha)^4 & 0 \\ 0 & 12/(\beta - \alpha)^2 \end{bmatrix}$

$\Lambda_{22} = -12/(\beta - \alpha)^2$, $L = 12/(\beta - \alpha)^2$ and R_n is asymptotically distributed as $\ell_1\chi_1^2 + \ell_2\chi_2^2$ where ℓ_1 and ℓ_2 are the eigenvalues of

$$\begin{bmatrix} 28.8/(\beta - \alpha)^4 & 0 \\ 0 & 12/(\beta - \alpha)^2 \end{bmatrix} \left(\begin{bmatrix} (\beta - \alpha)^4/72 & \\ 0 & (\beta - \alpha)^2/12 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & (\beta - \alpha)^2/12 \end{bmatrix} \right)$$

$$= \begin{bmatrix} .4 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus $R_n \xrightarrow{D} .4\chi_1^2$. Similarly, W_n is asymptotically distributed as $c_1\chi_1^2$ where

$$C_1 = [(\beta - \alpha)^4/180].\ [72/(\beta - \alpha)^4] = .4$$

Hence, all three statistics have the same asymptotic distributions, viz., $.4\chi_1^2$.

3.4 Asymptotic Distributions of the Rao and Wald Statistics under Local Alternatives when the Model is Incorrect

In this section, corresponding to the result in Theorem 2.3.1, we will derive the asymptotic distributions of R_n and W_n when the model is incorrect under a sequence of alternatives of the form

$$\theta^* = \theta_o + \Delta/\sqrt{n}\ ,\ n = 1,\ 2,\ \ldots \qquad (3.4.1)$$

where $= (\Delta_1, \ldots, \Delta_k)$ is a vector of constants.

Theorem 3.4.1: Assume the regularity conditions C1'-C4' and C5 hold. Under the above local alternatives (3.4.1), the Rao statistic R_n is asymptotically distributed as a linear combination of i.i.d. noncentral chi-squared random variables:

$$R_n \xrightarrow{D} \sum_{i=1}^{k} \ell_i'\ \chi_i^2\ (1,\beta)$$

where $\chi_i^2\ (1,\beta)$, i=1, 2, ..., k are i.i.d. noncentral chi-squared random variables with 1 degree of freedom and noncentrality parameter $\beta = \Delta'C(\theta_o)\Delta$, and $\ell_1' \geq \ell_2' \geq \ldots \geq \ell_k'$ are the eigenvalues of the matrix $C(\theta_o)I^{-1}(\theta_o)$.

<u>Proof</u>: Expanding $R_n = V_{\hat{\theta}_o}' I^{-1}(\theta_o)V_{\theta_o}$ about θ^*, we obtain

$$R_n = V_{\theta^*}' I^{-1}(\theta^*)V_{\theta^*} + \sum_{j=1}^{k} (\theta_{oj} - \theta_j^*) \frac{\partial}{\partial\theta_j} (V_{\theta^*}' I^{-1}(\theta^*)V_{\theta^*})$$

$$+ \frac{1}{2} \sum_{i,j=1}^{k} (\theta_{oi} - \theta_i^*)(\theta_{oj} - \theta_j^*) \frac{\partial^2}{\partial\theta_i\partial\theta_j} (V_{\theta+}' I^{-1}(\theta_+) V_{\theta+})$$

where

$$\theta_+ = \theta^* + \lambda(\theta_o - \theta^*),\ 0 \leq \lambda \leq 1.$$

Using the same notations as before, let

$$\phi_\ell = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_\ell} \log f(X_i,\theta),\ \ell = 1, 2, \ldots, k,$$

$I^{-1}(\theta) = (I^{\ell m})_{kxk}$, and

$C(\theta) = (C_{\ell m})_{kxk}$, ℓ, $m = 1, 2 \ldots, k$.

So $V_\theta' I^{-1}(\theta) V_\theta = \sum_{\ell,m=1}^{k} I^{\ell m} \phi_\ell\phi_m$ and

$$\frac{\partial}{\partial\theta_j} (V_\theta' I^{-1}(\theta)V_\theta) = \sum_{\ell,m=1}^{k} \frac{\partial}{\partial\theta_j} (I^{\ell m} \phi_\ell\phi_m).$$

Now, $\frac{\partial}{\partial\theta_j} (I^{\ell m} \phi_\ell\phi_m) = \frac{\partial I^{\ell m}}{\partial\theta_j} \phi_\ell\phi_m + I^{\ell m} \frac{\partial\phi_\ell}{\partial\theta_j} \phi_m + I^{\ell m} \phi_\ell \frac{\partial\phi_m}{\partial\theta_j}$.

By assumption C3', $E[\frac{\partial}{\partial\theta_j} \log f(X,\theta^*)] = 0$, $j = 1, 2, \ldots, k$. So we have, by the Weak Law of Large Numbers,

$$\frac{1}{\sqrt{n}} \phi_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_j} \log f(X_i, \theta^*) \overset{P}{\to} 0.$$

In addition, $\sqrt{n}\ (\theta_{oj} - \theta_j^*) = \sqrt{n}\ (- \frac{\Delta j}{\sqrt{n}}) = -\Delta_j$ under the specified local alternatives and applying the Central Limit Theorem in conjunction with assumption C2' yields

$$\phi_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_j} \log f(X_i,\theta^*) \overset{D}{\to} N(0,\ C_{jj}\ (\theta^*)\ ),\ j=1,\ 2,\ \ldots\ ,\ k.$$

Now, we rewrite the second term in the expansion of R_n as

$$\sum_{j=1}^{k} \sqrt{n}\ (\theta_{oj} - \theta_j^*) \sum_{\ell,m=1}^{k} \{ \frac{\partial I^{\ell m}}{\partial\theta_j} (\frac{\phi_\ell}{\sqrt{n}})\phi_m + I^{\ell m} (\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_j})\phi_m$$

$$+ I^{\ell m}\ \phi_\ell (\frac{1}{\sqrt{n}}\ \frac{\partial\theta_\ell}{\partial\theta_j}) \}\ ,$$

and observe that

$$\frac{\partial I^{\ell m}}{\partial\theta_j} (\frac{1}{\sqrt{n}} \phi_\ell)\ \phi_m \overset{P}{\to} 0 \text{ for all } \ell,\ m,\ j, \text{ and}$$

$$I^{\ell m} (\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_j})\ \phi_m \overset{a.d.}{=} I^{\ell m}\ C_{j\ell}\phi_m.$$

Let $d_{jm} = \sum_{\ell=1}^{k} I^{\ell m}\ C_{j\ell} = (j,m)^{th}$ term in $C(\theta^*)\ I^{-1}(\theta^*)$ and note that

$\sum_{m=1}^{k} d_{jm}\ \phi_m$ is the jth component of $C(\theta^*)\ I^{-1}(\theta^*)\ V_{\theta^*}$. Thus

$$\sum_{j=1}^{k} \sqrt{n}\ (\theta_{oj} - \theta_j^*) \sum_{\ell,m=1}^{k} I^{\ell m} (\frac{1}{\sqrt{n}}\ \frac{\partial\phi_\ell}{\partial\theta_j})\ \phi_m \overset{a.d.}{=} -\Delta'\ C(\theta^*)\ I^{-1}(\theta^*)\ V_{\theta^*}$$

and similarly,

$$\sum_{\ell,m=1}^{k} I^{\ell m}\ \phi_\ell\ (\frac{1}{\sqrt{n}} \frac{\partial\phi_m}{\partial\theta_j}) \overset{a.d.}{=} \sum_{\ell,m=1}^{k} I^{\ell m}\ \phi_\ell C_{mj}$$

= jth component of $C(\theta^*)\, I^{-1}(\theta^*) V_{\theta^*}$. Thus, the second term in the expansion of R_n reduces asymptotically to $-2\Delta'\, C(\theta^*)\, I^{-1}(\theta^*) V_{\theta^*}$. Lastly, we consider the third term.

$$\frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(\frac{1}{n} V_\theta'\, I^{-1}(\theta) V_\theta \right)$$

$$= \sum_{\ell,m=1}^{k} \Big\{ \Big[\frac{\partial^2 I^{\ell m}}{\partial\theta_i \partial\theta_j} \cdot \frac{\phi_\ell}{\sqrt{n}} \cdot \frac{\phi_m}{\sqrt{n}} + \frac{\partial I^{\ell m}}{\partial\theta_j} \left(\frac{1}{\sqrt{n}} \frac{\partial\theta_\ell}{\partial\theta_i} \right) \frac{\phi_m}{\sqrt{n}} +$$

$$\frac{\partial I^{\ell m}}{\partial\theta_j} \left(\frac{\phi_\ell}{\sqrt{n}} \right) \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_m}{\partial\theta_i} \right) \Big] + \Big[\frac{\partial I^{\ell m}}{\partial\theta_i} \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_j} \right) \frac{\phi_m}{\sqrt{n}} + I^{\ell m} \left(\frac{1}{\sqrt{n}} \frac{\partial^2\phi_\ell}{\partial\theta_i \partial\theta_j} \right) \frac{\phi_m}{\sqrt{n}} +$$

$$I^{\ell m} \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_j} \right) \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_m}{\partial\theta_i} \right) \Big] + \Big[\frac{\partial I^{\ell m}}{\partial\theta_i} \left(\frac{\phi_\ell}{\sqrt{n}} \right) \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_m}{\partial\theta_j} \right) +$$

$$I^{\ell m} \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_i} \right) \left(\frac{1}{\sqrt{n}} \frac{\partial\phi_m}{\partial\theta_j} \right) + I^{\ell m} \left(\frac{\phi_\ell}{\sqrt{n}} \right) \left(\frac{1}{\sqrt{n}} \frac{\partial^2\phi_m}{\partial\theta_i \partial\theta_j} \right) \Big] \Big\} .$$

Noting that $(\theta_o - \theta^*) = -\frac{\Delta}{\sqrt{n}} \to 0$ as $n \to \infty$, we have $\theta_+ \to \theta^*$ as $n \to \infty$ and so

$$\frac{1}{\sqrt{n}} \phi_\ell \Big|_{\theta_+} \xrightarrow{P} 0, \quad \ell = 1, 2, \ldots, k \text{ and}$$

$$\frac{1}{\sqrt{n}} \frac{\partial\phi_\ell}{\partial\theta_i} \Big|_{\theta_+} \xrightarrow{P} C_{\ell i}, \quad \ell, i = 1, 2, \ldots, k.$$

By the same kind of argument as used in the proof of Theorem 3.1.1, we can show that all terms in $\frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(\frac{1}{n} V_\theta'\, I^{-1}(\theta)\, V_\theta \right)$ vanish asymptotically except for terms of the form

$$I^{\ell m} \{ (\frac{1}{\sqrt{n}} \frac{\partial \phi_\ell}{\partial \theta_i}) (\frac{1}{\sqrt{n}} \frac{\partial \phi_m}{\partial \theta_j}) + (\frac{1}{\sqrt{n}} \frac{\partial \phi_\ell}{\partial \theta_j}) (\frac{1}{\sqrt{n}} \frac{\partial \phi_m}{\partial \theta_i}) \}$$

which converge in probability to

$$I^{\ell m} \{ C_{\ell i} C_{mj} + C_{\ell j} C_{mi} \} \Big|_{\theta = \theta_o}$$

Since $\sum_{\ell,m=1}^{k} I^{\ell m} C_{\ell j} C_{mi} \Big|_{\theta = \theta_o}$

$$= \sum_{m=1}^{k} C_{mi} \sum_{\ell=1}^{k} I^{\ell m} C_{j\ell} \Big|_{\theta = \theta_o}$$

$$= \sum_{m=1}^{k} C_{\ell i} d_{jm} \Big|_{\theta = \theta_o}$$

$= (j,i)$th element in $C(\theta_o) I^{-1}(\theta_o) C(\theta_o)$,

and similarly, $\sum_{\ell,m=1}^{k} I^{\ell m} C_{\ell i} C_{mj} \Big|_{\theta = \theta_o}$ is the (i,j)th element in $C(\theta_o) I^{-1}(\theta_o) C(\theta_o)$, the third term reduces to

$\Delta' C(\theta_o) I^{-1}(\theta_o) C(\theta_o)\Delta$.

Thus, we have shown that

$$R_n \overset{a.d.}{=} V_{\theta o}' I^{-1}(\theta_o) V_{\theta o} - 2\Delta' C(\theta_o) I^{-1}(\theta_o) V_{\theta o} + \Delta' C(\theta_o) I^{-1}(\theta_o) C(\theta_o)\Delta$$

$$= (V_{\theta o} - C(\theta_o)\Delta)' I^{-1}(\theta_o) (V_{\theta o} - C(\theta_o)\Delta) \overset{a.d.}{=} Z' I^{-1}(\theta_o) Z$$

where Z is a k-variate normal random variable with mean $C(\theta_o)\Delta$ and covariance matrix $C(\theta_o)$. Finally, an application of Lemma 2.3.1 yields the desired result. //

Theorem 3.4.2: Assume the regularity conditions C1'-C4' hold. Under the local alternatives (3.4.1), W_n is asymptotically distributed as a linear combination of noncentral chi-squared random variables:

$$W_n \overset{D}{\to} \sum_{i=1}^{r} d_i' \chi_i^2 (1, \Delta' M^{-1} \Delta)$$

where $\chi_i^2 (1, \Delta' M^{-1} \Delta)$, i=1, 2, ... , r are i.i.d. noncentral chi-squared random variables with 1 degree of freedom and noncentrality parameter $\Delta' M^{-1} \Delta$ and $d_1' \geq d_2' \geq \ldots \geq d_r'$ are the eigenvalues of the matrix

$$M(\theta_o) \, [\, (I \ 0) \, I^{-1}(\theta_o) \, \binom{I}{0'} \,]^{-1} ,$$

where $M(\theta_o)$ is the upper r x r diagonal block of the matrix

$$\Lambda^{-1}(\theta_o) \, C(\theta_o) \, \Lambda^{-1}(\theta_o).$$

Proof: Partition $\theta^* = (\delta^{*\prime}, \gamma^{*\prime})'$, $\theta_o = (\delta_o', \gamma_o')'$ where δ^* and δ_o are r x 1 fixed vectors. Then

$$\sqrt{n} \, (\hat{\delta}_n - \delta_o) = \sqrt{n} \, (\hat{\delta}_n - \delta^*) + \sqrt{n} \, (\delta^* - \delta_o) = \sqrt{n} \, (\hat{\delta}_n - \delta^*) + \Delta .$$

By assumption C2: $\sqrt{n} \, (\hat{\delta}_n - \delta^*) \to N(0, M)$ where M is the upper r x r diagonal block of $\Lambda^{-1} C(\theta^*) \Lambda^{-1}$. In addition, $\theta^* \to \theta_o$ as $n \to \infty$. So

$$\sqrt{n} \, (\hat{\delta}_n - \delta_o) \overset{D}{\to} N(\Delta, M(\theta_o)) .$$

Also, $\hat{\theta}_n \overset{P}{\to} \theta_o$ and so $I^{-1}(\hat{\theta}_n) \overset{P}{\to} I^{-1}(\theta_o)$. Thus, under the local alternatives (3.4.1),

$$W_n \overset{a.d.}{=} Z' \, [\, (I \ 0) \, I^{-1}(\theta_o) \, \binom{I}{0'} \,]^{-1} \, Z \ ,$$

where Z is a r-variate normal random variable with mean Δ and covariance matrix $M(\theta_o)$. Upon application of Lemma 2.3.1, the result follows.

CHAPTER 4

PERFORMANCE OF THE LIKELIHOOD RATIO, RAO AND WALD STATISTICS UNDER MODEL MISSPECIFICATION

Having established the asymptotic results connected with the test statistics R_n and W_n, we are now in a position to examine and compare their performance against the likelihood ratio statistic $-2\log\lambda_n$.

In this chapter, we follow the approach adopted by Foutz and Srivastava (1977) in using the concept of Bahadur efficiency to compare the performance of these test statistics.

4.1 Asymptotic Relative Efficiencies of Test Statistics

In order to evaluate the performance of the three test statistics using the concept of Bahadur efficiency, we shall first state the definitions of a standard sequence, the level attained by a standard sequence and the approximate slope of a standard sequence as found in Foutz and Srivastava (1977). [In the following definitions, T_n stands for any sequence of statistics and not just $-2\log\lambda_n$.]

Definition 4.1.1: A sequence $T_n = T_n(X_1, X_2, \ldots, X_n)$ of measurable functions is called a standard sequence for testing the hypothesis $H_o: \theta \in \Theta_o$ in the model P if the following conditions are satisfied:

(i) Let P^n be the joint distribution of $X_1, X_2, \ldots, X_n$. For every $\theta \in \Theta_o$, there is a continuous distribution function $G_\theta(t)$ such that

$$\lim_{n\to\infty} P^n\{(x_1, \ldots, x_n): T_n(x_1, \ldots, x_n) \leq t\} = G_\theta(t)$$

for every t.

(ii) For every $\theta \in \Theta_o$, there is a constant $a(\theta)$, $0 < a < \infty$, such that

$$\log\{1-G_\theta(t)\} = -at\{1 + o_\theta(1)\}/2$$

where, as $t\to\infty$, $o_\theta(1) \to 0$ uniformly for $\theta \in \Theta_o$.

(iii) There exists a function $b(\theta)$ on $\Theta - \Theta_o$, with $0 < b < \infty$ such that for each $\theta \in \Theta - \Theta_o$,

$$\frac{T_n}{n} \to b(\theta) \quad \text{a.s. P.}$$

Definition 4.1.2: Let T_n be a standard sequence for testing H: $\theta \varepsilon \Theta_o$ in the model P, and let $G_\theta(t)$ be defined as in (ii) above. For any given data $X_1, X_2, \ldots, X_n$, the approximate level attained by $\{T_n\}$ for testing H in the model P is defined by

$$L_n(x_1, \ldots, x_n) = \sup \{1 - G_\theta(T(x_1, \ldots, x_n)) : \theta \varepsilon \Theta_o\}.$$

We now quote the following theorems from Foutz and Srivastava (1977):

Theorem 4.1.1 (Foutz and Srivastava): Under the conditions C1-C4, the approximate slope of $\{-2\log\lambda_n\}$ for testing H: $\theta \varepsilon \Theta_o$ in P is given by

$$S_{T_n}(\theta^*) = \inf\{\frac{1}{C_1(\theta)}, \theta \varepsilon \Theta_o\}.b_{T_n}(\theta^*)$$

for $\theta^* \varepsilon \Theta - \Theta_o$. The constants $\{c_1(\theta), \theta \varepsilon \Theta_o\}$ are the eigenvalues specified in Theorem 1.3.1, and $b_{T_n}(\theta^*)$ is the almost sure limit of $-2\log\lambda_n/n$ when the distribution of X is P for $\theta^* \varepsilon \Theta - \Theta_o$.

For completeness, we will also quote the following lemmas from Foutz and Srivastava (1977) that are used to prove Theorem 4.1.1:

Lemma 4.1.1: Let $\Phi(t)$ be the standard normal distribution, i.e.,

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

Then $\log\{1 - \Phi(t)\} = t^2 \{1 + O(1)\}/2$, where

$$t^{-2}\{2\log t + \log(2\pi)\} < O(1) < t^{-2}\{\log(2\pi) - 2\log(1/t + 1/t^3)\}$$

for every $t > 1$.

<u>Lemma 4.1.2</u>: Let $H_k(t)$ be the chi-squared distribution function with k degrees of freedom. Then

$$\log\{1 - H_k(t)\} = -t\ \{1 + 0_k(1)\}/2, \quad \text{where}$$

$$-2k\log(t/2)/t \le 0_k(1) \le \{\log(2\pi) - 2\log(t^{-1/2} - t^{-3/2})\}/t.$$

<u>Lemma 4.1.3</u>: Let $G(t; C_1, C_2, \ldots, C_i) = P(\sum_{i=1}^{r} C_i\chi_i^2 \le t)$ where $C_1 \ge C_2 \ge \ldots \ge C_r \ge 0$ and $\chi_1^2, \chi_2^2, \ldots, \chi_r^2$ are independent chi-squared random variables with 1 degree of freedom. Then

$$\log\{1 - G(t; C_1, C_2, \ldots, C_r)\} = -t\{1 + 0(1)\}/(2C_1), \quad \text{where}$$

$$-2C_1 r\ \log(t/(2C_1)\ )/t \le 0(1) \le C_1[\log(2\pi) - 2\log\{(C_1/t)^{1/2}$$

$$-(C_1/t)^{3/2}\}]/t$$

for $t > C_1$.

The proofs of Lemmas 4.1.1, 2 and 3 are given in Foutz and Srivastava (1977).

Since all the three test statistics R_n, W_n and $-2\log\lambda_n$ have the same form of asymptotic distributions, viz., a linear combination of i.i.d. chi-squared random variables with 1 degree of freedom, it is clear that the test statistics R_n and W_n must also be standard sequences for testing the hypothesis $H: \theta \varepsilon\ \Theta_o$ in the same way as $-2\log\lambda_n$. Also, their approximate slopes are similarly expressed as follows:

<u>Theorem 4.1.2</u>: Under the conditions C1-C5, the approximate slope of R_n for testing $H_o: \theta \varepsilon\ \Theta_o$ in P is given by

$$S_{R_n}(\theta^*) = \inf\ \{\ \frac{1}{\ell_1(\theta)}\ ,\ \theta \varepsilon\ \Theta_o\}\ .b_{R_n}\ (\theta^*)$$

for $\theta^* \varepsilon \Theta - \Theta_o$. The constants $\{\ell,(\theta), \theta \varepsilon \Theta_o$ are the eigenvalues specified in Theorem 3.1.1, and $b_{R_u}(\theta^*)$ is the almost sure limit of R_n/n when the distribution of X is P for $\theta^* \varepsilon \Theta - \Theta_o$.

Theorem 4.1.3: Under the conditions C1-C4, the approximate slope of $\{W_n\}$ for testing H_o: $\theta \varepsilon \Theta_o$ in P is given by

$$S_{Wn}(\theta^*) = \inf \{\frac{1}{d_1(\theta)}, \theta \varepsilon \Theta_o\} . b_{W_n}(\theta^*)$$

for $\theta^* \varepsilon \Theta - \Theta_o$. The constants $\{d_1(\theta), \theta \varepsilon \Theta_o\}$ are the eigenvalues specified in Theorem 3.2.1, and $b_{W_n}(\theta^*)$ is the almost sure limit of W_n/n when the distribution of X is P and for $\theta^* \varepsilon \Theta - \Theta_o$

The proofs of Theorem 4.1.2 and Theorem 4.1.3 follow exactly as that of Theorem 4.1.1.

Following Bahadur, the asymptotic relative efficiency of a standard sequence $\{U_n\}$ w.r.t. another standard sequence $\{V_n\}$ for testing the hypothesis H_o: $\theta \varepsilon \Theta_o$ against the alternative $\theta^* \varepsilon \Theta - \Theta_o$ is defined as

$$ARE(U_n, V_n) = S_{U_n}(\theta^*)/S_{V_n}(\theta^*) .$$

4.2 Examples

In this section, we shall consider some examples of model misspecification and compare the performance of R_n, W_n and T_n.

Example 4.2.1: Consider again the problem in Example 3.3.1. Based on the Poisson model with mean $\theta > 0$, we wish to test the hypothesis H_0: $\theta = \theta_o$. It is easy to show that the approximate slopes of T_n, W_n and R_n are

$$S_{T_n}(\theta^*) = [\theta_o/\mathrm{Var}(X)]\ 2\{(\theta_o - \theta^*) + \theta^*\log(\theta^*/\theta_o)\}\ ,$$

$$S_{W_n}(\theta^*) = \frac{\theta_o}{\mathrm{Var}(X)}\ \frac{(\theta^*-\theta_o)^2}{\theta_o}$$

respectively where $\theta^* \neq \theta_o$. Let ARE (T_1, T_2) represent the Bahadur asymptotic relative efficiency of T_1 with respect to T_2. Then

$$\mathrm{ARE}(W_n, T_n) = S_{W_n}(\theta^*)/S_{T_n}(\theta^*)$$

$$= \frac{(\theta^*-\theta_o)^2}{\theta^*}\ /\ \{2(\theta_o - \theta^*) + 2\theta^*(\log\theta^* - \log\theta_o)\}$$

$$= (r-1)^2/\{2r[1 - r + r\log r]\}$$

where $r = \theta^*/\theta_o$. Similarly,

$$\mathrm{ARE}(R_n, T_n) = S_{R_n}(\theta^*)/S_{T_n}(\theta^*)$$

$$= [(\theta^* - \theta_o)^2/\theta_o]/\ \{2[(\theta_o - \theta^*) + \theta^*\log(\theta^*/\theta_o)]\}$$

$$= (r - 1)^2/[2(1 - r + r\log r)] = r\ \mathrm{ARE}(W_n, T_n).$$

Clearly, $ARE(R_n, W_n) = r$. The following table lists the various asymptotic relative efficiencies for some values of $r = \theta^*/\theta_o$. From Table 1, we see that W_n performs better than both R_n and T_n when the true mean θ^* is less than the specified value θ_o. However, R_n fares better than both W_n and T_n when $\theta^* > \theta_o$ and does worse if $0 < \theta^* < \theta_o$. Thus, none of them can claim superiority over the others against all possible alternatives $\theta^* \in \Theta - \Theta_o$.

Remark: It should be cautioned here that the preceding comment is made while keeping in mind that any comparison made using approximate slopes is subject to the possibility of error. This is so because the approximate slope of a test statistic is not always a close indicator of the exact slope. For a deeper discussion of this, see Bahadur (1967).

Example 4.2.2: Consider the problem in Example 3.3.2. The chosen model is the family of normal distributions with variance $\theta_1 > 0$ and finite mean θ_2. We wish to test the hypothesis H_o: $\theta_1 = \sigma_o^2$. When the model is incorrect and the random variable has a uniform distribution on the interval (α, β), it can be shown directly that $\theta^* = (Var(X), E(X))' = ((\beta - \alpha)^2/12, (\beta + \alpha)/2)' = (\sigma_o^2, \theta_2^*)'$ if $\theta^* \in \Theta_o$. Also,

$$I(\theta^*) = \begin{pmatrix} \frac{1}{2\sigma_o^4} & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{pmatrix} ,$$

TABLE 1

ASYMPTOTIC RELATIVE EFFICIENCIES OF TEST STATISTICS
(POISSON MODEL, P SATISFIES CONDITIONS C1-C4)

$r = \theta^*/\theta_o$	ARE(R_n,W_n)	ARE(W_n,T_n)	ARE(R_n,T_n)
0.1	0.1	6.05	0.60
0.3	0.3	2.41	0.72
0.5	0.5	1.63	0.81
0.7	0.7	1.28	0.89
0.9	0.9	1.07	0.97
1.1	1.1	0.94	1.03
1.3	1.3	0.84	1.10
1.5	1.5	0.77	1.16
1.7	1.7	0.71	1.21
1.9	1.9	0.67	1.27
3.0	3.0	0.51	1.54
4.0	4.0	0.44	1.77
5.0	5.0	0.40	1.98
10.0	10.0	0.29	2.89
15.0	15.0	0.25	3.68

$$C(\theta^*) = \begin{bmatrix} \frac{0.2}{2\sigma_o^4} & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} -\frac{1}{2\sigma_o^4} & 0 \\ 0 & -\frac{1}{\sigma_o^2} \end{bmatrix},$$

so that $\Lambda' C^{-1}(\theta^*) \Lambda = \begin{bmatrix} \frac{1}{0.8\sigma_o^4} & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{bmatrix}$,

$L = \frac{1}{\sigma_o^2}$ and $M = 0.8\sigma_o^4$. Direct computations yield

$$R_n/n = (-1 + [(n-1)/n].\ S_n{}^2/\sigma_o^2\ )^2/2,$$

$$W_n/n = (S_n{}^2 - \sigma_o^2)^2/(2S_n{}^4), \text{ and}$$

$$T_n/n = -\log(S_n{}^2/\sigma_o^2) + [(n-1)/n]S_n{}^2/\sigma_o^2 - (n-1)/n,$$

so that as $n \to \infty$, we have

$$R_n/n \overset{P}{\to} (-1 + \text{Var}(X)\sigma_o^2)^2/2,$$

$$W_n/n \overset{P}{\to} (\text{Var}(X) - \sigma_o^2)^2/\{2[\text{Var}(X)]^2\}, \text{ and}$$

$$T_n/n \overset{P}{\to} -\log[\text{Var}(X)/\sigma_o^2] + \text{Var}(X)\sigma_o^2 - 1,$$

where under $\theta^* \varepsilon\ \Theta - \Theta_o$, $\text{Var}(X) \neq \sigma_o^2$. Also, all three statistics have the same limiting distributions with $C_1 = 0.4$. Thus, their approximate slopes are

$$S_{R_n}(\theta^*) = (1/0.4)[-1 + \mathrm{Var}(X)/\sigma_o^2]^2/2$$

$$= [-1 + \mathrm{Var}(X)/\sigma_o^2]^2/0.8,$$

$$S_{W_n}(\theta^*) = (1/0.4)[\mathrm{Var}(X)/\sigma_o^2 - 1]^2/\{2(\mathrm{Var}(X)/\sigma_o^2]^2\} ,$$

and $S_{T_n}(\theta^*) = (1/0.4)\{ -\log[\mathrm{Var}(X)/\sigma_o^2] + \mathrm{Var}(X)/\sigma_o^2 - 1\}$, and the asymptotic relative efficiencies are

$$\mathrm{ARE}(R_n,T_n) = [-1 + \mathrm{Var}(X)\sigma_o^2\ ]^2/\{2[-\log[\mathrm{Var}(X)/\sigma_o^2] + \mathrm{Var}(X)\sigma_o^2 - 1]\} ,$$

$$\mathrm{ARE}(R_n,W_n) = [\mathrm{Var}(X)/\sigma_o^2]^2 \quad \text{and}$$

$$\mathrm{ARE}(W_n,T_n) = \mathrm{ARE}(R_n,T_n)/\mathrm{ARE}(R_n,W_n).$$

So the ARE's depend only on the ratio $\mathrm{Var}(X)/\sigma_o^2 = r$, say. We display in Table 2 some values of these ARE's for various values of r.

From Table 2, we see that for $r > 1$, R_n is to be preferred over W_n and T_n. For $r < 1$, W_n is to be preferred over R_n and T_n. However, if the true variance Var(X) is not known to be either larger or smaller than the hypothesized value σ_o^2, then a clear choice among the three test statistics cannot be made.

Example 4.2.3: Consider again the same problem as in Example 4.2.2 except that now the true distribution of X is a contaminated normal distribution of the form

$$Q_{\sigma^2,\mu,\gamma}(B) = .9P_{\sigma_1^2,\mu}(B) + .1P_{\sigma_2^2,\mu}(B)$$

TABLE 2

ASYMPTOTIC RELATIVE EFFICIENCIES OF TEST STATISTICS
(NORMAL MODEL, P - UNIFORM DIST. ON (α , β))

$r = Var(X)/\sigma_o^2$	$ARE(R_n, W_n)$	$ARE(W_n, T_n)$	$ARE(R_n, T_n)$
0.1	0.01	28.88	0.29
0.3	0.09	5.40	0.49
0.5	0.25	2.59	0.65
0.7	0.49	1.62	0.79
0.9	0.81	1.15	0.93
1.1	1.21	0.88	1.07
1.3	1.69	0.71	1.20
1.5	2.25	0.59	1.32
1.7	2.89	0.50	1.45
1.9	3.61	0.43	1.57
3.0	9.00	0.25	2.22
4.0	16.00	0.17	2.79
5.0	25.00	0.13	3.35
10.0	100.00	0.06	6.05
15.0	225.00	0.04	8.68

for measurable sets B where $P_{\sigma^2,\mu}$ denote the normal distribution with variance σ^2 and mean μ. (This example was considered in Foutz and Srivastava (1977).) The parameter $\gamma = \sigma_2^2/\sigma_1^2$ is assumed known. Since the model used is unchanged, the probability limits of $\frac{R_n}{n}$, $\frac{W_n}{n}$ and $\frac{T_n}{n}$ are the same as in Example 4.2.2. What is affected is the limiting distributions. By direct computation,

$$I(\theta^*) = \begin{pmatrix} \frac{1}{2\sigma_o^4} & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{pmatrix}$$

$$C(\ ^*) = \begin{pmatrix} \left[\frac{3(.9 + .1\gamma^2)}{(.9 + .1\gamma)^2} - 1\right] / \sigma_o^4 & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{pmatrix}$$

and $$\Lambda = \begin{pmatrix} -\frac{1}{2\sigma_o^4} & 0 \\ 0 & -\frac{1}{\sigma_o^2} \end{pmatrix}$$

So $$\Lambda' C^{-1}(\theta^*)\Lambda = \begin{pmatrix} \frac{1}{[3(.9+.1\gamma^2)/(.9+.1\gamma)^2-1]\sigma_o^4} & 0 \\ 0 & \frac{1}{\sigma_o^2} \end{pmatrix}$$

$L = \frac{1}{\sigma_o^2}$ and $M = \left(\frac{3(.9+.1\gamma^2)}{(.9+.1\gamma)^2} - 1\right) \sigma_o^4$

It is then straightforward to see that all three statistics have the same limiting distributions with the largest characteristic root

$$C_1 = \frac{1.5(.9 + .1\gamma^2) - .5(.9 + .1\gamma)^2}{(.9 + .1\gamma)^2}$$

Hence, the form of the ARE's remain the same as those computed in Example 4.2.2 because the approximate slope is the product of $\frac{1}{C_1}$ and the corresponding probability limit. It is thus possible to compute the ARE's for various values of γ. First, we note that under the contaminated normal distribution,

$$\mathrm{Var}(X) = (.9 + .1\gamma)\sigma_1^2 ,$$

So that for a fixed value of γ, the ARE's depend only on the ratio $\sigma_1^2/\sigma_o^2 = r$, say, as follows:

$$\mathrm{ARE}(R_n,T_n) = \frac{.5\{(.9 + .1\gamma)\sigma_1^2/\sigma_o^2 - 1\}^2}{\{-\log[(.9 + .1\gamma)\sigma_1^2/\sigma_o^2] + (.9 + .1\gamma)\sigma_1^2/\sigma_o^2 - 1\}} ,$$

$$\mathrm{ARE}(R_n,W_n) = (.9 + .1\gamma)^2 \; (\sigma_1^2/\sigma_o^2)^2 \quad \text{and}$$

$$\mathrm{ARE}(W_n,T_n) = \mathrm{ARE}(R_n,T_n)/\mathrm{ARE}(R_n,W_n).$$

Table 3 displays some values of the ARE's for various values of γ and σ_1^2/σ_o^2 .

Note: If $\gamma = 1$, i.e., there is no contamination, then the ARE values computed under the uniform (α, β) distribution are also those that belong to the present case of the normal distribution without contamination; i.e., when there is correct model specification. In other words, under the normal model, the three test statistics are able to discern departures from the null hypothesis in terms of variance while remaining relatively insensitive to model misspecification.

Also, the ARE's here exhibit the same trends as in previous examples; i.e., in general, R_n is more efficient than both T_n and W_n when σ_1^2/σ_o^2 is large ($\geq$ 1) and W_n is more efficient than both R_n and T_n for small values of σ_1^2/σ_o^2 ($\leq$ 1).

TABLE 3

ASYMPTOTIC RELATIVE EFFICIENCIES OF TEST STATISTICS
(NORMAL MODEL, P - CONTAMINATED NORMAL)

Table 3a. $\gamma = 2$

$r = \sigma_1^2/\sigma_o^2$	$ARE(R_n,W_n)$	$ARE(W_n,T_n)$	$ARE(R_n,T_n)$
0.1	0.01	24.85	0.30
0.3	0.11	4.70	0.51
0.5	0.30	2.26	0.68
0.7	0.59	1.42	0.84
0.9	0.98	1.03	1.01
1.1	1.46	0.78	1.14
1.3	2.04	0.63	1.28
1.5	2.72	0.52	1.42
1.7	3.50	0.44	1.55
1.9	4.37	0.39	1.68
3.0	10.89	0.22	2.39
4.0	19.36	0.16	3.01
5.0	30.25	0.12	3.62
10.0	121.00	0.05	6.58
15.0	272.25	0.03	9.46

* $\gamma = \sigma_2^2/\sigma_1^2$

(Table 3 Continued)

Table 3b. $\gamma = 3$

$r = \sigma_1^2/\sigma_o^2$	ARE(R_n, W_n)	ARE(W_n, T_n)	ARE(R_n, T_n)
0.1	0.01	21.68	0.31
0.3	0.13	4.14	0.54
0.5	0.36	2.01	0.72
0.7	0.71	1.26	0.89
0.9	1.17	0.90	1.05
1.1	1.74	0.69	1.21
1.3	2.43	0.56	1.36
1.5	3.24	0.47	1.51
1.7	4.16	0.40	1.65
1.9	5.20	0.35	1.80
3.0	12.96	0.20	2.56
4.0	23.04	0.14	3.24
5.0	36.00	0.11	3.90
10.0	144.00	0.05	7.11
15.0	324.00	0.03	10.24

(Table 3 Continued)

Table 3c. $\gamma = 10$

$r = \sigma_1^2/\sigma_o^2$	ARE(R_n,W_n)	ARE(W_n,T_n)	ARE(R_n,T_n)
0.1	0.04	10.68	0.39
0.3	0.32	2.15	0.70
0.5	0.90	1.07	0.97
0.7	1.77	0.69	1.21
0.9	2.92	0.50	1.45
1.1	4.37	0.39	1.68
1.3	6.10	0.31	1.91
1.5	8.12	0.26	2.13
1.7	10.43	0.23	2.35
1.9	13.03	0.20	2.57
3.0	32.49	0.11	3.73
4.0	57.76	0.08	4.76
5.0	90.25	0.06	5.78
10.0	361.00	0.03	10.76
15.0	812.25	0.02	15.66

CHAPTER 5

NONOPTIMALITY OF THE RAO STATISTIC UNDER CORRECT MODEL

Bahadur (1965) showed that under certain regularity conditions, there exists an upper bound to the exact slope of any standard sequence and, further, that the exact slope of the likelihood ratio test statistic achieves this upper bound.

In this chapter, we will show that whenever the Rao statistic possesses an exact slope, this slope does not attain this upper bound and, thus, it cannot be an optimal sequence.

5.1 Preliminaries

In previous chapters, we have examined the asymptotic properties of the Rao and Wald statistics in conjunction with that of the likelihood ratio statistic under model misspecification. The examples in Chapter 4 show clearly that when the model is incorrect, there is no overall superiority of one method over the others. In most cases, it is quite reasonable to expect that a statistic will perform better than the others only for a subset of the class of alternatives and actually do worse outside this subset.

However, a different situation exists when the model is correct. Bahadur (1965) established a certain kind of optimality property for the likelihood ratio test statistic. To be precise, he showed that under certain regularity conditions, the exact slope (which corresponds closely to the approximate slope and is defined in terms of the exact finite sample distribution of the sequence) of any standard sequence is bounded from above by a constant which is a function of the Kullback-Leibler information number. (In fact, as shown in Raghavachari (1970), this upper bound exists without any regularity conditions whatsoever.) Further, he also showed that under suitable regularity conditions, the exact slope of the likelihood ratio statistic achieves this upper bound. This at once establishes the likelihood ratio test statistic as an optimal sequence.

5.2 Nonoptimality of the Rao Statistic Under Correct Model

We shall first state one of the results contained in Bahadur (1965). For any $\theta \in \Theta$ and $\theta_o \in \Theta_o$, let

$$I(\theta, \theta_o) = -\int \log[f(x, \theta_o)/f(x,\theta)]\, dP_\theta$$

be the Kullback-Leibler information measure of $f(x,\theta)$ w.r.t. $f(x,\theta_o)$ and define

$$J(\theta) = \inf \{ I(\theta, \theta_o): \theta_o \in \Theta_o \}.$$

We assume that for each $\theta \in \Theta - \Theta_o$ and $\theta_o \in \Theta_o$ such that $I(\theta, \theta_o)$ there exists a $t = t(\theta, \theta_o) > 0$ such that

$$\int [f(x, \theta)/f(x, \theta_o)]^t \, dP_\theta < \infty$$

Let T_n be a measurable function of $\dot{S} = (x_1, x_2, \ldots)$ that depends on S only through $x_1, x_2, \ldots, x_n$. For each θ, define

$$F_n(t, \theta) = P_\theta (T_n(\dot{S}) < t)$$

$$G_n(t) = \inf \{ F_n(t, \theta): \theta \in \Theta_o \} \qquad \text{and}$$

$$L_n(S) = 1 - G_n (T_n(\dot{S})) .$$

<u>Theorem 5.2.1 (Bahadur)</u>: For each $\theta \in \Theta - \Theta_o$,

$$\liminf_{n \to \infty} \frac{1}{n} \log L_n(\dot{S}) \geq -J(\theta)$$

with probability one when θ obtains.

<u>Remark</u>: If P_θ admits a density w.r.t. P_{θ_o} , say, $dP_\theta = f(x)dP_{\theta_o}$, let

$$I^*(\theta, \theta_o) = E_\theta[\log f(x)];$$

otherwise, let $I^*(\theta, \theta_o) = \infty$. Also let

$$J^*(\theta) = \inf \{I^*(\theta, \theta_o): \theta_o \in \Theta_o\} .$$

Then, without assuming any regularity conditions, Raghavachavi (1970) showed that

$$\liminf_{n \to \infty} \frac{1}{n} \log L_n(S) \geq -J^*(\theta)$$

with probability one when Θ obtains.

In the following, we will consider the case when is a subset of the real line. Let $X_1, X_2, \ldots , X_n$ be i.i.d. observations from the distribution P_θ. We wish to test the hypothesis H_o: $\theta = \theta_o$ where θ_o is a specified constant. In this case, the Kullback-Leibler information number is

$$I(\theta, \theta_o) = \int \log[f(x, \theta)/f(x, \theta_o)] \, dP_\theta .$$

This represents the optimal slope of any test statistic; i.e., for any statistic T_n,

$$\liminf_{n \to \infty} \frac{1}{n} \log P_{\theta_o} (T_n > T_n(x_1, \ldots , x_n)) \geq -I(\theta, \theta_o)$$

Now consider the Rao statistic

$$R_n = \frac{1}{n} [\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i, \theta_o)]^2/I(\theta_o)$$

where $I(\theta_o) = E_{\theta_o} [\frac{\partial}{\partial\theta} \log f(X, \theta_o)]^2$. For convenience and ease of computation, we will use the equivalent statistic $S_n = \frac{R_n}{\sqrt{n}}$.

To compute the exact slope of S_n, we need to evaluate

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} \{S_n > S_n(x_1, \ldots , x_n)\} .$$

When the right situation exists, this can be accomplished through the following two theorems due to Bahadur:

Theorem 5.2.2 (Bahadur): Suppose K_n satisfies the following two conditions:

(a) For $\theta \varepsilon \Theta - \Theta_o$, $\frac{K_n}{\sqrt{n}} \to b(\theta)$ a.s. $P_{\hat{\Theta}}$, $-\infty < b(\theta) < \infty$.

(b) There exists an open interval I containing { $b(\theta):\theta \varepsilon \Theta - \Theta_o$ and a function g continuous on I such that

$$\lim_{n\to\infty} -\frac{2}{n} \log [1 - G_n(\sqrt{n}\ t)] = g(t),\ t \varepsilon I.$$

Then for $\theta \varepsilon \Theta - \Theta_o$

$$-\frac{2}{n} \log [1 - G_n (K_n(x_1, \ldots , x_n))] \to g(b(\theta)) \text{ a.s. } P_\theta .$$

Theorem 5.2.3 (Bahadur): Let Y_1, Y_2 ... denote a sequence of i.i.d. observations of Y, an extended real-valued random variable much that $P(-\infty \leq Y < \infty) = 1$. Let u be a real variable and let the function f be defined by

$$\exp [-f(u)] = \inf \{e^{-tu}\phi(t):\ \ t \geq 0 \}$$

where $\phi(t) = E(e^{ty})$. Also let $\{u_n\}_{n=1}^{\infty}$ be a sequence of constants such that

$$\lim_{n\to\infty} u_n = u,\ -\infty < u < \infty ,$$

and assume P (Y > u) > 0. Then

$$\lim_{n\to\infty} \frac{1}{n} \log P(Y_1 + \ldots + Y_n \geq n\, u_n) = -f(u).$$

A proof of Theorem 5.2.2 can be foundin Serfling (1980) while the reader is referred to the article of Bahadur (1971) for a proof of Theorem 5.2.3.

Let us now apply these two theorems to this problem. By the Strong Law of Large Numbers,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) \to C(\theta) \quad \text{a.s. } P_\theta$$

where

$$C(\theta) = \begin{cases} 0, \text{ if } \theta = \theta_o \\ E_\theta[\frac{\partial}{\partial\theta} \log f(X,\theta_o)], \text{ if } \theta \neq \theta_o \end{cases} .$$

Thus, it follows that

$$\frac{1}{\sqrt{n}} S_n = \frac{R_n}{n} = [\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o)]^2/I(\theta_o)$$

and this converges a.s. P_θ to $[C(\theta)]^2/I(\theta_o)$. Consider the following large-deviation probability of S_n under the null hypothesis. For $t > 0$,

$$P_{\theta_o}(S_n > \sqrt{n}\, t) = P_{\theta_o}(\frac{R_n}{n} > t)$$

$$= P_{\theta_o}\{|\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o)| > \sqrt{I(\theta_o)t}\}$$

$$= P_{\theta_o}\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) > \sqrt{I(\theta_o)t}\}$$

$$+ P_{\theta_o}\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) < -\sqrt{I(\theta_o)t}\} .$$

Let us first consider the first term. Assuming that

$$P_{\theta_o}\{\frac{\partial}{\partial\theta}\log f(X,\theta_o) > \sqrt{I(\theta_o)t}\} > 0,$$

and letting $u_n = \sqrt{I(\theta_o)t} = u$, it follows from Theorem 5.2.3 that

$$\lim_{n\to\infty}\frac{1}{n}\log P_{\theta_o}\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i,\theta_o) > n\sqrt{I(\theta_o)t}\} = -h(\sqrt{I(\theta_o)t})$$

where

$$\exp[-h(\sqrt{I(\theta_o)t})] = \inf\{e^{-\delta\sqrt{I(\theta_o)t}}\phi(\delta) : \delta \geq 0\}$$

$$= \inf\{E_{\theta_o}[e^{\delta(Y - \sqrt{I(\theta_o)t})}] : \delta \geq 0\} \text{ and } Y = \frac{\partial}{\partial\theta}\log f(X,\theta_o).$$

Example 5.2.1: Consider the following special case. Suppose P_θ represents the Poisson distribution with mean $\theta > 0$. Then

$$f(x,\theta) = e^{-x}\theta^x/x!,\ x = 0, 1, \ldots$$

$$Y = \frac{\partial}{\partial\theta}\log f(X,\theta_o) = -1 + X/\theta_o \quad \text{and}$$

$$I(\theta_o) = E_{\theta o}(-1 + X/_{\theta o})^2 = 1/\theta_o.$$

Thus $\exp[-h(\sqrt{I(\theta_o)t})] = \inf\{e^{-\delta\sqrt{t/\theta_o}}\, e^{[-\delta + \theta_o(e^{\delta/\theta_o} - 1)]} : \delta \geq 0\}$

so that $-h(\sqrt{I(\theta_o)t}) = \inf\{-\delta(1 + \sqrt{t/\theta_o}) + \theta_o(e^{\delta/\theta_o} - 1) : \delta \geq 0\}$.

We can show directly that the infemum is achieved at

$$\delta^* = \theta_o \log(1 + \sqrt{t/\theta_o}) > 0 \text{ for } t > 0 \quad \text{and so}$$

$$-h(\sqrt{I(\theta_o)t}) = \sqrt{t\theta_o} - \theta_o(1 + \sqrt{t/\theta_o})\log(1 + \sqrt{t/\theta_o}) = g_1(t).$$

Also, $C(\theta) = \begin{cases} 0, \text{ if } \theta = \theta_o \\ \frac{\theta}{\theta_o} - 1, \text{ if } \theta \neq \theta_o \end{cases}$

$$= \frac{\theta}{\theta_o} - 1 \quad \text{for all } \theta,$$

so that $\frac{S_n}{\sqrt{n}} = \frac{1}{n} R_n \to (\theta - \theta_o)^2/\theta_o$ a.s. P_θ. Now, an application of Theorem 5.2.3 yields

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} \{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) > \sqrt{I(\theta_o)t} \}$$

$$= \sqrt{t\theta_o} - \theta_o (1 + \sqrt{t/\theta_o}) \log (1 + \sqrt{t/\theta_o}) = g_1(t), \quad t > 0.$$

Similarly, we can also show (by considering $-Y = -\frac{\partial}{\partial\theta} \log f(X, \theta_o)$ instead of Y) that

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} \{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) < -\sqrt{I(\theta_o)t} \}$$

$$= -\sqrt{t\theta_o} - \theta_o(1 - \sqrt{t/\theta_o}) \log (1 - \sqrt{t/\theta_o}) = g_2(t), \text{ for } 0 < t < \theta_o.$$

In view of the domain of definition of $g_2(t)$, i.e. the interval $(0, \theta_o)$, we shall restrict the set of alternatives to $\{\theta: 0 < (\theta - \theta_o)^2/\theta_o < \theta_o\}$ so that $\Theta = \{\theta: 0 < \theta < 2\theta_o\}$ and the actual testing problem becomes testing

$$H_o: \theta = \theta_o \quad \text{vs.} \quad H_1: 0 < \theta < \theta_o \text{ or } \theta_o < \theta < 2\theta_o.$$

With this restriction, then $\{b(\theta) = (\theta - \theta_o)^2/\theta_o: \theta \in \Theta\}$ is contained in the open interval (θ, θ_o) in which both $g_1(t)$ and $g_2(t)$ are continuous.

<u>Remark</u>: If we wish to simplify the discussion, Θ could simply be taken as the interval $(0, \theta_o)$ and the testing problem is then one-sided, i.e., we will be testing

$$H_o: \ \theta = \theta_o \ \text{ vs. } \ H_1: \ 0 < \theta < \theta_o .$$

Now, $\frac{1}{n} \log P_{\theta_o} [S_n > \sqrt{n}\, t]$

$$= \frac{1}{n} \log [P_{\theta_o} \{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) > \sqrt{I(\theta_o)t} \ \}$$

$$+ P_{\theta_o} \{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) < - \sqrt{I(\theta_o)t} \ \}].$$

Since for any sets A_1 and A_2,

$$\log[P_\theta(A_1) + P_\theta(A_2)] \geq \max \{\log P_\theta(A_1), \log P_\theta(A_2)\} ,$$

we have

$$\limsup_{n \to \infty} \frac{1}{n} \log P_{\theta_o} (S_n > \sqrt{n}\, t) \geq \liminf_{n \to \infty} \frac{1}{n} \log P_{\theta_o} (S_n > \sqrt{n}\, t)$$

$$\geq \max \{ \lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} [\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) > \sqrt{I(\theta_o)t} \ ],$$

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} [\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta_o) < - \sqrt{I(\theta_o)t} \ ] \}$$

$$= \max\{\sqrt{t\theta_o} - \theta_o(1 + \sqrt{t/\theta_o} \log (1 + \sqrt{t/\theta_o}),$$

$$- \sqrt{t\theta_o} - \theta_o(1 - \sqrt{t/\theta_o}) \log (1 - \sqrt{t/\theta_o})\} = \underline{L} (t), \text{ say.}$$

This implies that whenever the limit

$$\lim_{n\to\infty} \frac{1}{n} \log P_{\theta_o} (S_n > \sqrt{n}\, t)$$

exists, it cannot be less than $\underline{L}$ (t). It then follows from Theorem 5.2.2 that the exact slope of S_n, when it exists, cannot exceed $-\underline{L}[(\theta-\theta_o)^2/\theta_o]$ where

$$\underline{L}[(\theta-\theta_o)^2/\theta_o] = \max\{|\theta-\theta_o|-\theta_o(1+|\theta-\theta_o|/\theta_o)\log[1+|\theta-\theta_o|/\theta_o],$$
$$-|\theta-\theta_o|-\theta_o(1-|\theta-\theta_o|/\theta_o)\log[1-|\theta-\theta_o|/\theta_o]\}$$

$$= \begin{cases} \max\{(\theta-\theta_o)-\theta\log(\theta/\theta_o), \\ \quad (\theta_o-\theta)+(2\theta_o-\theta)\log[\theta_o/(2\theta_o-\theta)]\}, & \theta_o < \theta < 2\theta_o \\ \max\{(\theta_o-\theta)-(2\theta_o-\theta)\log[(2\theta_o-\theta)/\theta_o], \\ \quad (\theta-\theta_o)+2\log(\theta_o/\theta)\}, & 0 < \theta < \theta_o \end{cases}$$

$$= \max\{(\theta-\theta_o)-\theta\log(\theta/\theta_o), (\theta_o-\theta)+(2\theta_o-\theta)\log[\theta_o/(2\theta_o-\theta)]\}.$$

Now, the optimal slope of any standard sequence is given by

$$I(\theta,\theta_o) = \sum_{x=o}^{\infty} \log\left[\frac{e^{-\theta}\theta^x/x!}{e^{-\theta_o}\theta_o^x/x!}\right] \frac{e^{-\theta}\theta^x}{x!} = \theta_o - \theta + \theta\log(\theta/\theta_o);$$

i.e., for any standard sequence $\{T_n\}$,

$$\liminf_{n\to\infty} \frac{1}{n} P_{\theta o}\{T_n > T_n(x_1, \ldots, x_n)\} \geq -I(\theta,\theta_o).$$

However, $\underline{L}[(\theta-\theta_o)^2/\theta_o]$

$$= \max\{-I(\theta,\theta_o), (\theta_o-\theta)+(2\theta_o-\theta)\log[\theta_o/(2\theta_o-\theta)]\}.$$

Thus, we are done if we can show that for some $\theta \varepsilon (0, 2\theta_o)$,

$$(\theta_o-\theta)+(2\theta_o-\theta)\log[\theta_o/(2\theta_o-\theta)] > -I(\theta,\theta_o).$$

By letting $q = \theta/\theta_o$, the problem is reduced to showing that there exists a real number q, $0 < q < 2$, such that

$$2(1-q) > q \log [q/(2-q)] + 2 \log (2-q).$$

Utilizing the inequality $\log x \leq x-1$ for $x > 0$, the r.h.s. of the above equation becomes

$$q \log q + (2-q) \log (2-q) \leq q^2 - q + (2-q)(2-q-1) = 2(q-1) < 2(1-q)^2$$

for $o < q < 1$.

This shows that S_n does not attain the optimal slope for $0 < \theta < \theta_o$.

5.3 The 1-Parameter Exponential Model

Let us consider the more general setting where P belongs to the 1-parameter exponential family defined by the density

$$f(x,\theta) = K(\theta)\, e^{\theta T(x)}, \quad -\infty < x < \infty$$

We will restrict ourselves to those members of this family that satisfy the regularity conditions C1-C5. In this case, θ^* is defined by

$$E\left[\frac{\partial}{\partial\theta} \log f(X,\theta^*) \right] = \frac{K'(\theta^*)}{K(\theta^*)} + E[T(X)] = 0,$$

where $K'(\theta) = \frac{\partial}{\partial\theta} K(\theta)$. This equation can be rewritten as

$$\log[K(\theta^*)] = -\theta^* E[T(X)].$$

Similarly, the m.l.e. $\hat{\theta}_n$ (when it exists) is given by the equation

$$\log K(\hat{\theta}_n) = -\hat{\theta}_n \overline{T(X)}$$

where $\overline{T(X)} = \frac{1}{n} \sum_{i=1}^{n} T(X_i)$. We shall restrict ourselves to the problem of testing the simple hypothesis, H_o: $\theta = \theta_o$, where θ_o is a specified constant. By direct computation, we deduce

$$I(\theta^*) = \{E\,[T(X)]\}^2 - \{2E\,[T(X)]\, E_\theta\,[T(X)] + E_\theta[T(X)]^2\}\big|_{\theta^*},$$

$C(\theta^*) = \operatorname{Var}\,[T(X)]$, and $\Lambda(\theta^*) = K''(\theta^*)/K(\theta^*) - \{E[T(X)]\}^2$

Thus, $\Lambda^{-1}C(\theta^*)\Lambda^{-1}$

$$= \frac{E[T(X)]^2 - \{E[T(X)]\}^2}{(K''(\theta^*)/K(\theta^*) - \{E[T(X)]\}^2)^2}$$

= M, the matrix defined in Theorem 1.3.1. Also, the other matrix $W = -\Lambda$ and so under the 'null' hypothesis $\theta_o = \theta^*$, we have

$$MW = \frac{E[T(X)]^2 - \{E[T(X)]\}^2}{[\{E[T(X)]\}^2 - K''(\theta_o)/K(\theta_o)]}$$

Similarly, it is easy to show that

$$C(\theta^*)I^{-1}(\theta^*) = \frac{E[T(X)]^2 - \{E[T(X)]\}^2}{\{E[T(X)]\}^2 - 2E[T(X)]E_\theta[T(X)]|_{\theta_o} + E[T(X)]^2|_{\theta_o}}$$

and $MI(\theta^*)$

$$= (E[T(X)]^2 - \{E[T(X)]\}^2) (\{E[T(X)]\}^2 - 2E[T(X)]E_\theta[T(X)]|_{\theta_o}$$

$$+ E_\theta[T(X)]^2|_{\theta_o})/(K''(\theta_o)/K(\theta_o) - \{E[T(X)]\}^2)^2 .$$

Thus, all three statistics will have the same asymptotic distributions if $\{E[T(X)]\}^2 - K''(\theta_o)/K(\theta_o)$

$$= \{E[T(X)]\}^2 - 2E[T(X)]\, E_\theta\, [T(X)]|_{\theta_o} + E_\theta[T(X)]^2|_{\theta_o} ,$$

which reduces to

$$-K''(\theta_o)/K(\theta_o) = -2E[T(X)]\, E_\theta\, [T(X)]|_{\theta_o} + E_\theta\, [T(X)]^2|_{\theta_o} .$$

Remarks: (a) Let P_θ be the Poisson distribution with mean $\theta > 0$. The density function is

$$f(\chi, \theta) = e^{-\theta}\theta^\chi/\chi!, \ \chi = 0, 1, \ldots$$

and the above condition reduces to $E(X) = \theta_o$.
Since $\theta^* = E(X)$, this is satisfied under the 'null' hypothesis $\theta^* - \theta_o$.

(b) Let P_θ be the binomial distribution with parameters n and θ,

$0 < \theta < 1$. Consider n fixed. The density function is

$$f(\chi, \theta) = \binom{n}{\chi} \theta^{\chi}(1 - \theta)^{n-\chi}, \chi = 0, 1, \ldots, n$$

and the condition becomes $E(X) = n\theta_o$.

Again, since $\theta^* = E(X)/n$, thus under the 'null' hypothesis $\theta^* = \theta_o$, the condition is satisfied and all three statistics have the same asymptotic distributions.

5.4 An Example of 2-Parameter Exponential Family - The Normal Model

Let P_θ represent the normal distribution specified by the density

$$F(\chi,\ \theta) = (2\pi\sigma^2)^{-1/2} \exp\{-(\chi - \mu)^2/(2\sigma^2)\},\ -\infty < \chi < \infty$$

where $\theta = (\mu,\ \sigma^2)'$. Suppose the null hypothesis of interest is

H_o: $\theta = \theta_o$, where $\theta_o = (\mu_o\ \sigma_o^2)'$ is a constant vector. It is easy to show that $\mu^* - E(X)$ and $^2{*} = Var(X)$. These are the almost sure limits of the m.l.e. $\hat{\theta}_n = (\overline{X}\ \ S_n^2)'$, where $\overline{X}$ is the sample mean and S_n^2 is the sample variance $\sum_{i=1}^{n} (\chi_i - \bar{\chi})^2/n$. Also,

$$I(\theta^*) = \begin{bmatrix} 1/Var(X) & 0 \\ 0 & 0.5/[Var(X)]^2 \end{bmatrix},$$

$$C(\theta^*) = \begin{bmatrix} 1/Var(X) & -\dfrac{E(X - \mu^*)^3}{2[Var(X)]^3} \\ -\dfrac{E(X - \mu^*)^3}{2[Var(X)]^3} & 0.5/[Var(X)]^2 \end{bmatrix},$$

$$\text{and } \Lambda = \begin{bmatrix} -1/Var(X) & 0 \\ 0 & -0.5/[Var(X)]^2 \end{bmatrix},$$

Thus, we have

$$M = \Lambda^{-1}C(\theta^*)\Lambda^{-1} = \begin{bmatrix} Var(X) & -E(X - \mu^*)^3 \\ -E(X - \mu^*)^3 & 2[Var(X)]^2 \end{bmatrix},$$

$$W = -\Lambda = \begin{pmatrix} 1/\mathrm{Var}(X) & 0 \\ 0 & 0.5/[\mathrm{Var}(X)]^2 \end{pmatrix} , \text{ and}$$

$$MW = \begin{pmatrix} 1 & -E(X-\mu^*)^3/\{2[\mathrm{Var}(X)]^2 \\ -D(X-\mu^*)^3/\mathrm{Var}(X) & 1 \end{pmatrix} .$$

The eigenvalues of this matrix are

$$C_1 = 1 + \frac{E(X-\mu^*)^3}{\mathrm{Var}(X)\ \sqrt{2\mathrm{Var}(X)}} \quad \text{and} \quad C_2 = 1 - \frac{E(X-\mu^*)^3}{\mathrm{Var}(X)\ \sqrt{2\mathrm{Var}(X)}} ,$$

So that under $\theta^* = \theta_o$, $-2\ \log\lambda_n \xrightarrow{D} C_1\chi_1^2 + C_2\chi_2^2$

where χ_1^2 and χ_2^2 are independent chi-squared random variables with 1 degree of freedom. Similarly, it is easy to verify that the corresponding matrix for the Rao statistic R_n is

$$\begin{pmatrix} 1 & -E(X-\mu^*)^3/\mathrm{Var}(X) \\ -E(X-\mu^*)^3/\{2[\mathrm{Var}(X)]^2\} & 1 \end{pmatrix} ,$$

and the matrix for W_n is

$$\begin{pmatrix} 1 & -E(X-\mu^*)^3/\{2[\mathrm{Var}(X)]^2\} \\ -E(X-\mu^*)^3/\mathrm{Var}(X) & 1 \end{pmatrix} .$$

Since these possess the same eigenvalues as MW, we deduce that all three statistics have the same asymptotic distributions under the incorrect model and if $\theta^* = \theta_o$; i.e., if $E(X) = \mu_o$ and $Var(X) = \sigma_o^2$. This is true regardless of the underlying distribution P as long as conditions C1-C5 hold.

5.5 Concluding Remarks

We have shown in this work that under model misspecification, none of the three competing procedures can claim outright superiority over the others (at least not with regard to the criterion of approximate Bahadur efficiency).

However, when the model is correct, the likelihood ratio test statistic is better than the Rao statistic since the former is an optimal sequence and the latter is not. It would be desirable if the same kind of statement can be made concerning the Wald statistic. Unfortunately, the calculation of large deviation probabilities in this case proved to be intractable.

A closely related subject concerns tests of model misspecification. White (1980) has recently proposed such a test. However, its properties and operating characteristics have not been studied. More research in this direction is needed and perhaps, other tests of model misspecification would arise in the course of such research.

BIBLIOGRAPHY

Bahadur, R.R. (1965). An Optimal Property of the Likelihood Ratio Statistic, Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, 1, pp. 13-26.

Bahadur, R.R. (1967). Rates of Convergence of Estimates and Test Statistics, Annals of Mathematical Statistics, 38, pp. 303-324.

Bahadur, R.R. (1971). Some Limit Theorems in Statistics, Regional Conference Series in Applied Mathematics published by the Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Bhattacharya, R.N. and Rao, R. Ranga (1976). Normal Approximation and Asymptotic Expansions, Wiley, New York.

Fisher, R.A. (1921). On the Mathematical Foundations of Theoretical Statistics, Philosophical Transactions of the Royal Society of London, Series A, 222, pp. 309-368.

Foutz, Robert V. and Srivastava, R.C. (1974). The Asymptotic Distribution of the Likelihood Ratio when the Model is Incorrect, Technical Report No. 110, Department of Statistics, The Ohio State University, Columbus, Ohio.

Foutz, Robert V. and Srivastava, R.C. (1977). The Performance of the Likelihood Ratio Test when the Model is Incorrect, Annals of Statistics, 5, pp. 1183-1194.

Foutz, Robert V. and Srivastava, R.C. (1978). The Asymptotic Distribution of the Likelihood Ratio when the Model is Incorrect, The Canadian Journal of Statistics, 6, No. 2, pp. 273-279.

Huber, P.J. (1965). The Behavior of the Maximum Likelihood Estimates under Non-standard Conditions, Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, 1, pp. 221-233.

Rao, C.R. (1947). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation, Proceedings of the Cambridge Philosophical Society, 44, pp. 50-57.

Rao, C.R. (1963). Criteria of Estimation in Large Samples, Sankhya, Series A, 25, pp. 189-206.

Raghavachari, M. (1970). On a Theorem of Bahadur on the Rate of Convergence of Test Statistics, Annals of Mathematical Statistics, 41, pp. 1695-1699.

Serfling, Robert J. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.

Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large, Transactions of American Mathematical Society, 54, pp. 426-482.

White, H. (1980). Maximum Likelihood Estimation of Misspecified Models: I & II, University of Rochester Department of Economics Discussion Paper, Rochester, New York.

Wolfowitz, J. (1965). Asymptotic Efficiency of the Maximum Likelihood Estimator, Theory of Probability and its Applications, 40, pp. 247-260.