ON LIKELIHOOD RATIO TESTS OF ONE-SIDED HYPOTHESES IN GENERALIZED LINEAR MODELS WITH CANONICAL LINKS

by

Mervyn J. Silvapulle

School of Agriculture, La Trobe University
Bundoora, Australia

(Prepared while a Visiting School at the Department of Biostatistics, Univ. of N.C., Chapel Hill, NC)

Institute of Mimeo Series No. 1877

May 1990

ON LIKELIHOOD RATIO TESTS OF ONE-SIDED HYPOTHESES IN GENERALIZED LINEAR MODELS WITH CANONICAL LINKS

Mervyn J. Silvapulle*

Department of Biostatistics, University of North Carolina at Chapel Hill.

* Address for corespondence : School of Agriculture, La Trobe University, Bundoora, Australia 3083.

ABSTRACT

For generalized linear models with multivariate response and natural link functions, likelihood ratio test of one-sided hypothesis on the regression parameter is considered under rather general conditions. The null-asymptotic distribution of the test statistic turns out to be chi-bar squared. The extension of the above results to include quasi-likellihood ratio test to incorporate over-dispersion when the response is univariate is also discussed. A simple example illustrates the application of the main result.

Keywords and phrases : asymptotic distribution; chi-bar squared distribution; logistic regression; quasi-likelihood;

Running Head : Generalized Linear Models.

1. INTRODUCTION

Let us consider a set of n observations $(y_1, Z_1), \ldots, (y_n, Z_n)$, where y_i is a q-dimensional random vector, usually referred to as the dependent variable, and Z_i is a $p \times q$ matrix of associated explanatory variables, i= 1, ..., n. We assume that $y_1, \ldots, y_n$ are independent and that the explanatory variables are nonstochastic. We shall also assume that, for i=1, ..., n, the density of y_i is

$$f(y_i; \theta_i) = c(y_i)\exp\{\theta_i^t y_i - b(\theta_i)\} \tag{1.1}$$

and that

$$\theta_i = Z_i^t \beta_o \tag{1.2}$$

where β_o is a $p \times 1$ vector of unknown parameters and $b(\cdot)$ is some function. Clearly this belongs to the class of generalized linear models with canonical link function (1.2) (see, Nelder and Wedderburn (1972), McCullagh and Nelder (1983, Chapter 2), and Fahrmeir and Kaufmann (1985)). Usually the parameter β_o is estimated by $\hat{\beta}_n$, the maximum likelihood estimator.

Likelihood ratio test of hypotheses of the form $R\beta_o = 0$ against $R\beta_o \neq 0$, where R is a given matrix is discussed in McCullagh (1983). Often in practice, the possible directions of the effects of explanatory variables on y are known, and hence the alternative hypothesis tend to be one-sided, for example $R\beta_o \geq 0$. It appears that this problem, for the generalized linear models, has not been investigated adequately in the literature. Obviously, when the null and alternative hypotheses are of the form $H_1 : R\beta_o = 0$ and $H_2 : R\beta_o \geq 0$ respectively, it is desirable to use a test that makes use of the information contained in the alternative hypothesis than to simply apply a test that is designed for testing $R\beta_o = 0$ against $R\beta_o \neq 0$. In this paper, we derive the asymptotic distribution of the likelihood ratio statistic under the null hypothesis, when the alternative hypothesis is one-sided. The hypotheses are stated in a rather general form. An example in Section 4 illustrates the application of the main results.

2. PRELIMINARIES

We will use some of the results in Fahrmeir and Kaufmann (1985), and therefore, it is convenient to use some of their notations as well. So, let $l_n(\beta)$ be the log-likelihood corresponding to (1.1) and (1.2), $F_n(\beta) = \text{cov}\{(\partial/\partial\beta)l_n(\beta)\} = -(\partial^2/\partial\beta\partial\beta^t)l_n(\beta)$, and $F_n = F_n(\beta_o)$. Thus, F_n is the information matrix. In general, F_n is positive definite, and therefore we may write $F_n = F_n^{1/2}F_n^{t/2}$, where $F_n^{t/2} = (F_n^{1/2})^t$ with t denoting the transpose as usual. Let us now define two regularity conditions:

(D): The smallest eigen value of $F_n \to \infty$ as $n \to \infty$.

(N): For every $\delta > 0$, Max $\{ \| F_n^{-1/2}F_n(\beta)F_n^{-t/2} - I \| : \beta \in N_n(\delta)\} \to 0$ as $n \to \infty$, where $N_n(\delta) = \{\beta : \| (\beta - \beta_o)^t F_n^{1/2} \| \leq \delta\}$.

The above conditions (D) and (N) are the same as (D) and (N) in Fahrmeir and Kaufmann (1985).

The following formulation of the hypotheses is essentially the same as in Kodde and Palm (1986). Let h be a $k \times 1$ vector function of β where $k \leq p$. Assume that $(\partial/\partial\beta)h(\beta)$ is continuous at β_o, and that rank$\{(\partial/\partial\beta)h(\beta_o)\} = k$. Let h be divided into two subvectors, h_1 and h_2, consisting of the first k_1 and the remaining $(k - k_1)$ elements of h respectively. Now, let us define the null and alternative hypotheses as follows.

$$H_0 : h(\beta_o) = 0 \qquad \text{and} \qquad H_1 : h(\beta_o) \neq 0,\ h_2(\beta_o) \geq 0. \tag{2.1}$$

Let C_o and C_1 be the closures of the null and alternative parameter spaces respectively, and assume that they are convex. For most practical applications, it suffices to restrict h to linear functions only.

For large n, since $(\hat{\beta}_n - \beta_o) \sim N(O, F_n^{-1})$ we have $\{h(\hat{\beta}_n) - h(\beta_o)\} \sim N(O, \Pi_n)$, where $\Pi_n = \Pi_n(\beta_o)$, and $\Pi_n(\beta) = (\partial/\partial\beta^t)h(\beta)F_n^{-1}(\beta)\{(\partial/\partial\beta)h(\beta)\}$. Let Π_n be partitioned to conform with the partitioning of h into h_1 and h_2; let us write the two rows of the resulting matrix as $[\Pi_{11n}\ \ \Pi_{12n}]$ and $[\Pi_{21n}\ \ \Pi_{22n}]$.

For any given covariance matrix A of order $j \times j$, let w(j, i, A) be the probability that exactly i components of a j-dimensional N(O, A) random variable are positive. For discussions on the interpretation and computation of w(j, i, A) see Gourieroux et.al. (1982), Wolak (1987) and Shapiro (1988).

3. THE MAIN RESULTS.

Let LR be the likelihood ratio statistic for testing H_0 against H_1. So, we have

$$LR = 2\,[\sup\{l_n(\beta) : \beta \in C_1\} - \sup\{l_n(\beta) : \beta \in C_0\}]. \tag{3.1}$$

For $c > 0$, let

$$\eta_n(\beta, c) = \sum_{i=0}^{k-k_1} pr\{\chi^2(k_1 + i) \geq c\}\; w\{k - k_1, i, \Pi_{22n}(\beta)\}. \tag{3.2}$$

The main theoretical result of this paper is the following:

Theorem. Assume that conditions (D) and (N) are satisfied. Then, for a fixed $c > 0$, we have

$$pr(LR \geq c \mid H_0) - \eta_n(\beta_0, c) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

The proof of the theorem is given in Section 5. It will be clear that , for the above testing problem, the Wald statistic and the likelihood ratio statistic have the same asymptotic distribution under the null hypothesis. However, this does not mean that the properties of the two statistics are the same. In contrast to the likeliood ratio statistic, the Wald statistic is not invariant to the choice of the function h which defines the null and alternative hypotheses. Further, Hauck and Donner (1977) show that, in logistic regression, the Wald statistic may behave rather erratically compared to the likelihood ratio statistic. For more discussions on this issue see Væth (1985) and Phillips and Park (1988).

Often in empirical studies involving the generalized linear models, the data exhibit over-dispersion. To incorporate this phenomenon, we may adopt the quasi-likelihood approach (see, McCullagh and Nelder (1983, Chapter 8)). For more discussions on other appoaches to dealing with over-dispersion/extra-variation, see Wilson (1989) and the references therein. Assume that the response variable is of one dimension; that is, y_i is a scalar. Let μ_i and $v(\mu_i)$ be the mean and variance corresponding to (1.1). In the quasi-likelihood approach, we do not assume that the response variable y_i follows a specific distribution such as (1.1). Let us assume that the mean and variance of y_i are respectively μ_i and $\sigma^2 v(\mu_i)$ for some $\sigma > 0$; $\sigma > 1$ correspond to over-dispersion. The quasi-likelihood $K(y_i, \mu_i)$ for the *i* th observation is a function satisfying $(\partial/\partial\mu_i) K(y_i, \mu_i) = \{(y_i - \mu_i)/v(\mu_i)\}$. One such function is $K(y_i, \mu_i) = \log f(y_i; \theta_i)$ since $v(\mu_i)$ is the variance function for $f(y_i; \theta_i)$ in (1.1) (see, Wedderburn (1974, Theorem 2)). Although, the quasi-likelihood approach allows a wide variety of choices for $v(\mu_i)$ in general, our results here are applicable only to those situations where $v(\mu_i)$ is the variance function corresponding to (1.1). However, these special cases are wide enough to incorporate a substantial proportion of practical situations.

The maximum quasi-likelihood estimator $\hat{\beta}_n$ of β_o is the value of β that maximizes the quasi-likelihood $\Sigma K\{y_i, \mu_i(\beta)\}$. Thus, $\hat{\beta}_n$ is the same as the maximum likelihood estimator corresponding to (1.1) and (1.2). Further, we have, $n^{1/2}(\hat{\beta}_n - \beta_o)$ is asymptotically $N(O, \sigma^2 F_n^{-1})$ (see, Fahrmeir and Kaufmann (1985)), and the quasi-likelihood ratio test statistic for testing H_o against H_1 is $\hat{\sigma}^{-2}LR$, where $\hat{\sigma}$ is a consistent estimator of σ and LR is the same as in (3.1); for a discussion on consistent estimation of σ, see McCullagh and Nelder (1983, Section 8.5). Now, we have the following from the above theorem.

Corollary Suppose that Conditions (D) and (N) are satisfied. Let $\hat{\sigma}$ be a consistent estimator of σ. Then , for $c > 0$, $pr(\hat{\sigma}^{-2}LR > c \mid H_o) - \eta_n(\beta_o, c) \rightarrow 0$ as $n \rightarrow \infty$.

It follows from the definition of $w\{k - k_1, i, \Pi_{22n}(\beta)\}$ and $\eta_n(\beta, c)$ that they are continuous in β. So, $\eta_n(\beta_o, c)$ may be estimated consistently by $\eta_n(\hat{\beta}_n, c)$. Thus, denoting the observed value of the test statistic $\hat{\sigma}^{-2}LR$ by c^*, $\eta_n(\hat{\beta}_n, c^*)$ provides an estimate of the p-value, $pr(\hat{\sigma}^{-2}LR > c^* \mid H_o)$. Therefore, the above theorem and corollary provide the basic asymptotic results for testing H_o against H_1 in (2.1) using likelihood ratio or quasi-likelihood ratio statistics. If $h = h_1$, then our results for the testing problem (2.1) reduce to the corresponding two-sided ones in McCullagh (1983).

Computation of the weights $w\{k - k_1, i, \Pi_{22n}(\beta)\}$ when $k - k_1$ is larger than 4, is not trivial. Therefore, prior to computing these weights, it would be desirable to obtain upper and lower bounds for the approximate p-value $\eta_n(\beta_o, c^*)$, where c^* is the observed value of the test statistic. By arguments similar to those for Lemma 6.1 and Theorem 6.2 of Perlman (1969), (see also Kodde and Palm (1986)) it may be shown that

$$\inf_A \sum_{i=0}^{k-k_1} pr\{\chi^2(k_1 + i) \geq c\}\, w\{k - k_1, i, A\} = 0.5[pr\{\chi^2(k_1) \geq c\} + pr\{\chi^2(k_1 + 1) \geq c\}] \quad (3.3)$$

$$= p_L(c), \text{ say}$$

$$\sup_A \sum_{i=0}^{k-k_1} pr\{\chi^2(k_1 + i) \geq c\}\ w\{k - k_1, i, A\} = 0.5[pr\{\chi^2(k - 1) \geq c\} + pr\{\chi^2(k) \geq c\}] \quad (3.4)$$

$$= p_U(c), \text{ say}$$

where inf and sup are taken over positive definite matrices, $\{A\}$, of order $(k - k_1)$. Thus, $p_L(c^*)$, and $p_U(c^*)$ are lower and upper bounds for the p-value. Therefore, in general, we need to compute the weights and then the approximate p-value $\eta_n(\hat{\beta}_n, c^*)$ only when the lower bound is 'small' and the upper bound is 'large'. Obviously, the above upper and lower bounds provides a reasonable 'safety-net', but it should however be recognized that these bounds could be quite far away from the true value.

4. AN EXAMPLE

In this section, an example is discussed to illustrate a practical role for the main theoretical results in the earlier sections. Therefore, no attempt is made to provide a complete statistical analysis of the data. This example is taken from Koch *et al* (1985, Section 3.5.5). Standard errors and p-values quoted in this section are all based on asymptotic results, and hence they are approximations only. The presence or absence of coronary artery desease, as based on angiographic evaluation as the standard diagnostic test, is analyzed in relationship to Age, Sex and an Electrocardiogram (ECG) screening outcome during exercise. On the basis of prior preliminary data and other relevant literature, the prevalence of coronary artery disease was expected to have a non-decreasing relationship with Age and severity of ECG, and males were expected to be at least as much prone to coronary artery disease as are the females. A test of whether or not at least one of these three factors had an increasing relationship with coronary artery disease was an object of the study.

Since the response variable is binary, let us consider the logistic regression model,

$$p(x_i, \beta) = \{ 1 + \exp(-\beta_1 - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4}) \}^{-1} \quad (4.1)$$

where, for the *i*th patient, $p(x_i, \beta)$ is the probability of current coronary artery disease, x_{i2} is the age in years, x_{i3} is 1 for males and 0 for females, and x_{i4} is an ECG score which takes the values 0, 1 or 2 according as ST-segment depression is less than 0.1, between 0.1 and 0.2 and above 0.2 respectively.

Thus, the null and the alternative hypotheses may be stated as

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{and} \quad H_1 : \beta_2, \beta_3, \beta_4 \geq 0\text{, and } (\beta_2, \beta_3, \beta_4) \neq 0. \tag{4.2}$$

In the notation of (2.1), we have $h(\beta) = h_2(\beta) = (\beta_2, \beta_3, \beta_4)^t$, $k = 3$ and $k_1 = 0$. Note that the formulation in (4.2) is different from the usual testing problem,

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{against} \quad H_2 : (\beta_2, \beta_3, \beta_4) \neq 0. \tag{4.3}$$

The set of data in Koch *et al* (1985), consists of 78 observations. To illustrate the main theoretical results in Section 3 and to highlight some of differences between one and two-sided tests, we found it convenient to use a smaller sample than that contained in Koch et al (1985). So we chose a subsample (not a random sample) of size 38, and for this subsample the results are given below.

For the logistic regression computations reported below, we used the statistical package SAS, but one could use almost any standard statistical package. The unrestricted estimate of $(\beta_2, \beta_3, \beta_4)$ together with their standard errors and individual p-values based on two-sided likelihood ratio statistics are 0.12 (se=0.07, p=0.08), 1.51 (se=0.90, p=0.09), and 0.91 (se=0.5, p=0.07) respectively. The standard errors are quoted for completeness. Since H_1 is one-sided, these p-values need to be divided by 2; the reason for this would be trivial had the p-values been based on t-statistics, but not so trivial for the above ones based on two-sided likelihood ratio statistics. The justification follows from the following two points: Let LR_j be the one-sided likelihood ratio statistic for $H_{jo} : \beta_j = 0$ against $H_{j1} : \beta_j > 0$; then (1) $pr(LR_j > c) \rightarrow 0.5pr\{\chi^2(1) > c\}$ as $n \rightarrow \infty$, for any $c > 0$ since $w(1, 0, a) = w(1, 1, a) = 0.5$ for any $a > 0$, and (2) the unrestricted maximum likelihood estimate of β_j falls in the one-sided parameter space for H_{j1}, and hence the one-sided and two-sided likelihood ratio statistics have the same numerical value. If we test for each coefficient separately, and adjust the overall p-value to account for the multiple tests, evidence in favour of H_1 in (4.2) is very weak.

Now, let us apply the one-sided likelihood ratio test to (4.2). To compute the likelihood ratio statistic, we need to compute $\sup \{l_n(\beta) : \beta \in C_1\}$ which is $\sup \{l_n(\beta) : \beta_2, \beta_3, \beta_4 \geq 0 \}$. Usually, to compute this, we need a constrained optimization routine similar to, for example, BCOAH in the IMSL (Version 10) package. However, for our example, the unrestricted estimates of β_2, β_3 and β_4 are positive, and hence $\sup \{l_n(\beta) : \beta_2, \beta_3, \beta_4 \geq 0 \} = \sup l_n(\beta) = 22.2$ (SAS and many other packages do provide the value of $-2\sup l_n(\beta)$). This is an important observation since in most practical situations, the unrestricted estimates do satisfy the inequalities in the alternative hypothesis, and hence there is no need to use a constrained minimization routine. It is easily seen that $\sup \{l_n(\beta) : H_o\} = \sup \{l_n(\beta) : \beta_2 = \beta_3 = \beta_4 = 0\} = [\, r \log(r/n) + (n-r) \log\{1-(r/n)\}\,] = -26.3$, where n and r are respectively the total number of patients and those with coronary artery disease. Therefore, by (3.1) the likelihood ratio statistic, $LR = 2(26.3 - 22.2) = 8.18$. By (3.3) and (3.4), lower and upper bounds for the p-value are

$$p_L = 0.5\mathrm{pr}\{\chi^2(0) > 8.18\} + 0.5\mathrm{pr}\{\chi^2(1) > 8.18\} = 0.002 \quad (4.4)$$

and $$p_U = 0.5\mathrm{pr}\{\chi^2(2) > 8.18\} + 0.5\mathrm{pr}\{\chi^2(3) > 8.18\} = 0.03 \quad (4.5)$$

respectively. Since the upper bound for the p-value given above is 'small' it is reasonable to conclude that there is sufficient evidence in favour of the alternative hypothesis. However, to find the strength of this evidence we need to compute the weights.

In view of (4.2), it is easily sean that $\Pi_{22n}(\hat{\beta}_n)$ is the submatrix of $F_n^{-1}(\hat{\beta}_n)$ obtained by deleting the first row and the first column. Since $(\hat{\beta}_n - \beta_o)$ is asymptotically $N\{O, \sigma^2 F_n^{-1}(\beta_o)\}$, most statistical packages provide $F_n^{-1}(\hat{\beta}_n)$ as an estimate of the covariance matrix of $\hat{\beta}_n$. The weights $w_i = w\{3, i, \Pi_{22n}(\hat{\beta}_n)\}$ for i = 0, 1, 2, 3 were computed rather easily by using the expressions derived by Kudo (1963) - these expressions are also given in Wolak (1987) in our notation. The computed values of w_i for i=0, 1, 2, and 3 respectively are 0.062, 0.297, 0.438 and 0.203. So, an estimate of the p-value for the one-sided likelihood ratio test is $\{ 0.297\,\mathrm{pr}(\chi_1^2 > 8.18) + 0.438\,\mathrm{pr}(\chi_2^2 > 8.18) + 0.203\,\mathrm{pr}(\chi_3^2 > 8.18) \} = 0.017$. Therefore, based on the one-sided test, we conclude that there is strong evidence in favour of H_1. On the other hand, if we carry out a likelihood ratio test of the two-sided hypothesis in (4.3), p-value = $\mathrm{pr}(\chi_3^2 > 8.18) = 0.043$. Thus, the two-sided test does not provide as much evidence in favour of H_1 as does the one-sided test .

In summary, we see that the implementation of the one-sided test is not all that difficult provided that the number of restrictions in the alternative hypothesis is not 'large'; as the number of restrictions increase, computation of the weights w(k, i, A) becomes non-trivial, but algorithms are available. However, lower and upper bounds for the p-value may be obtained quite easily. It is rather easy to see that, when the unrestricted estimates of the parameters satisfy the inequality restictions in the alternative hypothesis (frequently, this is the case), the p-value and its upper bound in (3.4) for the one-sided likelihood ratio test are *smaller* than the p-value for the corresponding two-sided test. Since the lower and upper bounds in (3.4) and (3.5) respectively are very easy to compute, perhaps we should consider computing them as a genaral rule whenever the alternative hypothesis is one-sided. Thus, it appears reasonable to conclude that the one-sided likelihood ratio test discussed in this paper would be useful whenever the alternative hypothesis is one-sided and the sample size is not large enough for the two-sided likelihood ratio test to detect the departure from the null hypothesis.

5. PROOF OF THE THEOREM

Let us first state and prove a lemma. The proof of the Theorem in Section 3 follows easily from this lemma and the main results in Kodde and Palm (1986).

Lemma. Assume that Conditions (D) and (N) are satisfied. Let C be a closed convex set in the p-dimensional Euclidean space, $\beta_o \in C$, and $\bar{\beta}_n$ and $\tilde{\beta}_n$ be defined by $\| F_n^{t/2}(\bar{\beta}_n - \hat{\beta}_n) \| = \inf \{\| F_n^{t/2}(\beta - \hat{\beta}_n) \| : \beta \in C\}$ and $l_n(\tilde{\beta}_n) = \sup \{l_n(\beta) : \beta \in C\}$, respectively. Then, we have

(i) $\| F_n^{t/2}(\bar{\beta}_n - \beta_o) \|^2 = O_p(1)$; (ii) $F_n^{t/2}(\tilde{\beta}_n - \beta_o) = O_p(1)$;

(iii) $l_n(\beta) = l_n(\hat{\beta}_n) - 2^{-1}\| F_n^{t/2}(\beta - \hat{\beta}_n) \|^2 + U_n(\beta)$,

where $\sup\{ |U_n(\beta)| : \beta \in N_n(\delta)\} \to 0$ in probability;

(iv) $l_n(\tilde{\beta}_n) = l_n(\hat{\beta}_n) - 2^{-1}\sup \{\| F_n^{t/2}(\beta - \hat{\beta}_n)\|^2 : \beta \in C\} + o_p(1)$.

<u>Proof</u> (i) $\|F_n^{t/2}(\bar{\beta}_n - \beta_o)\|^2 \le \|F_n^{t/2}(\bar{\beta}_n - \hat{\beta}_n)\|^2 + \|F_n^{t/2}(\hat{\beta}_n - \beta_o)\|^2$

$\le 2\|F_n^{t/2}(\hat{\beta}_n - \beta_o)\|^2$, since $\beta_o \in C$.

$= O_p(1)$.

(ii) By Theorem 3 of Fahrmeir and Kaufmann (1985), $F_n^{t/2}(\hat{\beta}_n - \beta_o)$ is asymptotically normal. Further, since $\{l_n(\beta) - l_n(\beta_o)\}$ is concave, given $\eta > 0$, there exists $\delta > 0$ such that

$$pr[\ \sup\{\ l_n(\beta) - l_n(\beta_o) : \beta \in \partial N_n(\delta) \cap C\ \} < 0\ ] > (1 - \eta)$$

for large n, where ∂ stands for boundary. Now, since $N_n(\delta) \cap C$ is a convex set and $l_n(\beta) - l_n(\beta_o)$ is a concave function of β, we may verify that $pr\{\tilde{\beta}_n \in N_n(\delta)\} > (1 - \eta)$, for large n. To see this, assume that the event E occured, where E = ' $\sup\{\ l_n(\beta) - l_n(\beta_o) : \beta \in \partial N_n(\delta) \cap C\ \} < 0$ '. Suppose that $\tilde{\beta}_n \in N_n(\delta)$. Let the line segment $[\beta_o, \tilde{\beta}_n]$ intersect $\partial N_n(\delta) \cap C$ at β^*. Then, we have $l_n(\beta^*) < l_n(\beta_o)$ since E occured, and $l_n(\tilde{\beta}_n) > l_n(\beta_o)$ by the definition of $\tilde{\beta}_n$. This is a contradiction since $l_n(\beta)$ is concave along the line segment $[\beta_o, \tilde{\beta}_n]$. Therefore, $\tilde{\beta}_n \in N_n(\delta)$.

(iii) Given $\beta \in C$, we can find $\bar{\beta}$ lying between β and $\hat{\beta}_n$ such that

$$l_n(\beta) = l_n(\hat{\beta}) - 2^{-1}\|F_n^{t/2}(\beta - \hat{\beta}_n)\|^2 - U_n(\beta).$$

where $U_n(\beta) = 2^{-1}(\beta - \hat{\beta}_n)^t F_n^{1/2}\ \{F_n^{-1/2} F_n(\bar{\beta}) F_n^{-t/2} - I\} F_n^{t/2}(\beta - \hat{\beta})$. By Condition (N), $\sup\{\ |U_n(\beta)| : \beta \in N_n(\delta)\} \to 0$ in probability.

(iv) This part follows from (i), (ii) and (iii).Now, if the null hypothesis is true then by the above lemma and the fact that $\{l_n(\beta) - l_n(\beta_o)\}$ is concave, we have LR = D + $o_p(1)$, where

$$D = \inf\{\ \|F_n^{t/2}(\beta - \hat{\beta}_n)\|^2 : h(\beta) = 0\} - \inf\{\|F_n^{t/2}(\beta - \hat{\beta}_n)\|^2 :\ h_2(\beta) \geq 0\}.$$

Clearly, D is a Wald-type statistic. Kodde and Palm (1986) refer to D as a Distance statistic and D-statistic. Now, the proof of our theorem follows from Kodde and Palm (1986, Case 1).

Acknowledgment : I am grateful to Professor G.G.Koch for very helpful discussions on some aspects of the example in Section 4.

REFERENCES

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann.of Statist.* 13, 342-368.

Gourieroux, C., Holly, A. and Montfort, A. (1982). Likelihood ratio, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*, 50, 63-80.

Hauck, W.W. and Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Jour. Amer. Statist. Assoc.*, 72, 851-853.

Kodde, D.A. and Palm, F.C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54, 1243-1248.

Koch, G.G., Imrey, P.B., Singer, J.M., Atkinson, S.S., and Stokes, M.E. (1985). *Analysis of Categorical data.* Les presses de L'Universite de Montreal.

McCullagh, P (1983). Quasi-likelhood functions. *Ann. Statist, 11, 59-67.*

McCullagh, P. and Nelder, J.A. (1983). *Generalized linear models.* Chapman and Hall, London.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Jour. Roy. Statist. Soc. Ser. A*, 135, 370-384.

Perlman, M.D. (1969). One-sided problems in multivariate anakysis. *Ann. Math. Statist.*, 40, 549-569.

Phillips, P.C.B. and Park, J.Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica*, 56, 1065 - 1083.

Shapiro, A (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Inter.Statist. Rev.*, 56, 49-62.

Væth, M. (1985). On the use of Wald's test in exponential families. *Inter. Statist. Rev.*, 53, 199-214.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. *Biometrika*, 61, 439-447.

Wilson, J. R. (1989). Chi-square tests for overdispersion with multiparameter estimates. *Appl. Statist*, 38, 441-453

Wolak, F.A. (1987). An exact test for multiple inequality and equality constraints in the linear model. *Jour. Amer. Statist. Assoc.*, 82, 782-793.