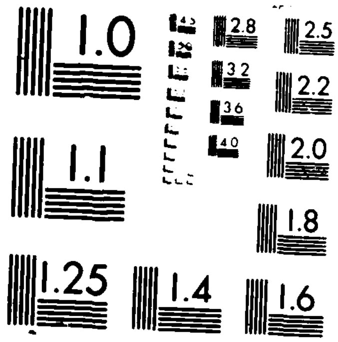
AD-A186 032   VARIABLE SELECTION IN LOGISTIC REGRESSION(U) PITTSBURGH   1/1
UNIV PA CENTER FOR MULTIVARIATE ANALYSIS
Z D BAI ET AL JUN 87 TR-87-23 AFOSR-TR-87-1074
UNCLASSIFIED F49620-85-C-0008   F/G 12/3   NL



ROCOPY RESOLUTION HART

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

(2)

AD-A186 032

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR- 87-1074	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Variable selection in logistic regression	5. TYPE OF REPORT & PERIOD COVERED ~~Technical~~ Journal - July 1987
	6. PERFORMING ORG. REPORT NUMBER 87-23
7. AUTHOR(s) Z. D. Bai, P. R. Krishnaiah and L. C. Zhao	8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR/NM Bldg 410 BAFB, DC 20332	12. REPORT DATE July 1987
	13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) AFOSR/NM Bolling AFB DC 20332-6448 Bldg 410	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Same

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC SELECTED OCT 14 1987 E

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Consistency, information theoretic criterion, logistic discrimination, logistic regression, maximum likelihood, model selection.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In many situations, we are interested in selection of important variables which are adequate for prediction under a logistic regression model. In this paper, some selection procedures based on the information theoretic criteria are proposed, and these procedures are proved to be strongly consistent.

DD FORM 1 JAN 73 1473

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR- 87-1074

VARIABLE SELECTION IN LOGISTIC REGRESSION

Z. D. Bai, P. R. Krishnaiah and L. C. Zhao

Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis

University of Pittsburgh

VARIABLE SELECTION IN LOGISTIC REGRESSION

Z. D. Bai, P. R. Krishnaiah and L. C. Zhao

Center for Multivariate Analysis
University of Pittsburgh

June 1987

Technical Report No. 87-23

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

Research sponsored by the Air Force Office of Scientific Research (AFSC) under contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

VARIABLE SELECTION IN LOGISTIC REGRESSION

Z. D. Bai, P. R. Krishnaiah and L. C. Zhao

Center for Multivariate Analysis
University of Pittsburgh

ABSTRACT

In many situations, we are interested in selection of important variables which are adequate for prediction under a logistic regression model. In this paper, some selection procedures based on the information theoretic criteria are proposed, and these procedures are proved to be strongly consistent.

AMS 1980 Subject Classifications. Primary 62H12, 62H15.

Key Words and Phrases: Consistency, information theoretic criterion, logistic discrimination, logistic regression, maximum likelihood, model selection.

1. INTRODUCTION

Logistic regression is the most used form of binary regression (see Berkson (1951), Cox (1970), and Efron (1975). The investigation of this aspect has had an important impact on disease diagnostics (refer to Gordon and Kannel (1968), Pregibon (1981) and, Stefanski and Carroll (1985)). One of the important aspects related to logistic regression is logistic discrimination (refer to J. A. Anderson (1982)).

The model to be considered is given by

$$P_r\{Y = 1|\underset{\sim}{X}\} = \{1 + \exp(-\beta_0-\beta_1 X^{(1)}- \dots -\beta_p X^{(p)})\}^{-1} \tag{1.1}$$

$$P_r\{Y = 0|\underset{\sim}{X}\} = 1 - P_r\{Y = 1|\underset{\sim}{X}\},$$

where $X' = (X^{(1)},\dots,X^{(p)})$ is a $p\times 1$ random vector.

In some situations there are many potential variables $X^{(i)}$'s. This may represent the experimenter's lack of knowledge, his caution, or both. One objective of the statistician must be to choose a set of good predictor variables from the set of possible variables. A similar problem may also be met in logistic discrimination. In this paper, we are interested in selection of important variables that are adequate for prediction in the regression model (1.1). Using an information theoretic criterion, we propose some selection procedures which are strongly consistent.

In Section 2, the above problem is formulated, and the main methods and results are stated. Some lemmas are introduced in Section 3, and the Section 4 is devoted to the proof of the theorems.

2. PROBLEM AND MAIN RESULTS

Let $(\underset{\sim}{X},Y)$ be a random vector such that $\underset{\sim}{X}$ is a p-vector and Y is Bernouli variable with

$$P_r\{Y = 1|\underset{\sim}{X}\} = p(\underset{\sim}{Z}'\underset{\sim}{\beta}) \triangleq \{1 + \exp(-\underset{\sim}{Z}'\underset{\sim}{\beta})\}^{-1}, \tag{2.1}$$

where $\underset{\sim}{\beta}' = (\beta_0,\ldots,\beta_p)$, $\underset{\sim}{Z}' = (1,\underset{\sim}{X}') = (1,X^{(1)},\ldots,X^{(p)})$. Assume that F, the distribution of X, satisfies the following conditions:

(i) If $\underset{\sim}{\beta} \neq \underset{\sim}{\gamma}$, then

$$F\{X: p(\underset{\sim}{Z}'\underset{\sim}{\beta}) \neq p(\underset{\sim}{Z}'\underset{\sim}{\gamma})\} > 0. \tag{2.2}$$

(ii) $E(\underset{\sim}{X}'X) < \infty$.

Put $A = \{0,1,\ldots,p\}$. It is easily seen that, there exist a unique subset B_0 of A such that $\{0\}\subset B_0$, and, $i \in B_0$, $i \neq 0$ if and only if $\beta_i \neq 0$. Call B_0 the best subset of A. Note that if $\beta_i = 0$ for some $i \in A - \{0\}$, then Y is independent of $X^{(i)}$.

In this paper, we want to determine the best subset B_0 of A. To this end, suppose that $(\underset{\sim}{X}_1,Y_1),\ldots,(\underset{\sim}{X}_n,Y_n)$ are iid. observations of $(\underset{\sim}{X},Y)$. A step-wise selection method based on testing a series of hypotheses is proposed by J. A. Anderson (1982, pp.169-191). But it is difficult to seek for the conditional limit distribution of the test statistic for latter hypothesis after the former hypotheses was tested. In this paper, we propose a method based on the information theoretic criterion, and establish the strong consistency of this method under some mild conditions.

Let $\{0\} \subset B \subset A$. Write

$$M_B = \{\beta \in R^{p+1}: \beta_i = 0 \quad \text{for all } i \in A - B\}. \tag{2.3}$$

Let $L_n(\beta)$ be the likelihood function. Then

$$\log L_n(\beta) = \sum_{i=1}^{n} [Y_i \log p(z_i'\beta) + (1-Y_i) \log q(z_i'\beta)], \tag{2.4}$$

where $q(.) = 1 - p(.)$, $z_i' = (1, X_i') = (1, X_i^{(1)}, \ldots, X_i^{(p)})$. Put

$$G_n(B) = \sup_{\beta \in M_B} \log L_n(\beta), \tag{2.5}$$

and

$$I_n(B) = G_n(B) - \#(B)\, C_n, \tag{2.6}$$

where C_n satisfies the following conditions:

$$\lim_{n\to\infty} C_n/n = 0 \quad \text{and} \quad \lim_{n\to\infty} C_n/\log\log n = \infty, \tag{2.7}$$

Choose $\hat{B}$ such that $\{0\} \subset \hat{B} \subset A$ and

$$I_n(\hat{B}) = \max_{B:\{0\}\subset B \subset A} I_n(B), \tag{2.8}$$

and use $\hat{B}$ as an estimate of the best subset B_0 of A. We have the following

THEOREM 2.1. Under the condition (2.2), $\hat{B}$ is a strongly constant estimate of the best subset B_0 of $A = \{0,1,2,\ldots,p\}$.

Note that the above consistency means that with probability one for n large, $\hat{B}$ coincides with the best subset of A.

For simplicity of calculation, we can use another alternative method. To this end, put

$$A^{(i)} = A - \{i\}, \quad i = 1,2,\ldots,p.$$

There is one subset, either A or $A^{(i)}$, written as $B^{(i)}$, which satisfies

$$I_n(B^{(i)}) = \max\{I_n(A), \quad I_n(A^{(i)})\}, \quad i = 1,\ldots,p. \tag{2.9}$$

Put

$$\hat{B} = \bigcap_{i=1}^{p} B^{(i)},$$

then we can use $\hat{B}$ as an estimate of the best subset B_0 of A. In the same way, we have

THEOREM 2.2. Under the condition (2.2), $\hat{B}$ is strongly consistent

In the following sections, we will only give a proof for the theorem 2.1. The proof of the theorem 2.2 is similar and is omitted.

3. ASYMPOTOTIC EXPANSION OF SOME STATISTICS

Now we assume that $\beta = (\beta_0, \ldots, \beta_p)$ is the true parameter. Put

$$\frac{1}{n} \log L_n(\gamma) = \frac{1}{n} \sum_{i=1}^{n} [Y_i \log p(Z_i'\gamma) + (1-Y_i) \log q(Z_i'\gamma)] \tag{3.1}$$

$$\begin{aligned} H(\gamma) &= \int [P(Z'\beta) \log P(Z'\gamma) + q(Z'\beta) \log q(Z'\gamma)]dF \\ H_n(\gamma) &= \frac{1}{n} \sum_{i=1}^{n} [P(Z_i'\beta) \log P(Z_i'\gamma) + q(Z_i'\beta) \log q(Z_i'\gamma)]. \end{aligned} \tag{3.2}$$

Since $|\log p(u)| \leq 2 + |u|$, $|\log q(u)| \leq 2 + |u|$ for any real u, $H(\gamma)$ is finite for any $\gamma \in R^{p+1}$. For fixed β, functions $\frac{1}{n}\log L_n(\gamma)$, $H_n(\gamma)$ and $H(\gamma)$ are all concave in γ.

We need the following lemmas:

LEMMA 3.1. Let E be an open convex subset of R^p and let $f_1, f_2, \ldots,$ be a sequence of concave functions such that $\forall\, x \in E$, $f_n(x) \to f(x)$ as $n \to \infty$, where f is some real function on E. Then f is also concave and for all compact $D \subset E$,

$$\sup_{x \in D} |f_n(x) - f(x)| \to 0 \quad \text{as } n \to \infty.$$

LEMMA 3.2. Suppose that $\{f_n\}$ and f satisfy the conditions of the above lemma, and f has a unique maximum at $\hat{x} \in E$. Let $\hat{x}_n$ maximize f_n. Then $\hat{x}_n \to \hat{x}$ as $n \to \infty$.

For a proof of the above two lemmas, the reader is referred to Rockafellar (1970, Theorem 10.8), P. K. Anderson and R. D. Gill (1982, Theorem II.1, Corollary II.2).

LEMMA 3.3. Let $\hat{\beta}_n$ be a maximum likelihood estimate of β. If (i) of (2.2) and the following condition are satisfied:

$$E\|X\| < \infty \quad \text{with} \quad \|X\| = (X'X)^{1/2}. \tag{3.3}$$

Then,

$$\lim_{n\to\infty} \hat{\beta}_n = \beta \quad \text{a.s.,} \quad \lim \frac{1}{n}\log L_n(\hat{\beta}_n) = H(\beta) \quad \text{a.s.} \tag{3.4}$$

Proof. By Jensen's inequality,

$$H(\gamma) \le H(\beta) \quad \text{for any } \gamma \in R^{p+1}$$

and the equality holds if

$$F(X : P(Z'\gamma) = P(Z'\beta)) = 1$$

Thus, by the condition (i) of (2.2), $H(\gamma)$ has a unique maximum at β.

Now let $\hat{\beta}_n$ maximize $\frac{1}{n}\log L_n(\gamma)$. By (3.3) and the strong law of large numbers (Sl

$$\lim_{n\to\infty} \frac{1}{n}\log L_n(\gamma) = H(\gamma) \quad \text{a.s.} \tag{3.5}$$

for any $\gamma \in R^{p+1}$, and (3.4) follows from Lemmas 3.1 and (3.2).

Note that $\hat{\beta}_n$ satisfies the likelihood equation, that is

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i - p(Z_i'\hat{\beta}_n)]Z_i = 0. \tag{3.6}$$

We have the following lemma.

LEMMA 3.4. Suppose that the conditions of (2.2) are satisfied, then with probability one for n large,

$$\hat{\beta}_n - \beta = S(\beta)^{-1}(1+o(1))\frac{1}{n}\sum_{i=1}^{n}(Y_i - p(Z_i'\beta))Z_i \tag{3.7}$$

where

$$S(\gamma) = \int p(Z'\gamma)q(Z'\gamma)ZZ'dF > 0. \tag{3.8}$$

From this, $\hat{\beta}_n - \beta$ obeys the law of iterated logarithm, i.e.

$$\hat{\beta}_n - \beta = O\left(\sqrt{\frac{1}{n}\log\log n}\right) \qquad \text{a.s.} \tag{3.9}$$

Proof. At first we show that $S(\gamma) > 0$. Otherwise, there exists some constant $(p+1)$ - vector $C \neq 0$ such that $E(C'Z)^2\, p(Z'\gamma)q(Z'\gamma) = 0$, i.e.,

$$F\{X : C'Z = 0\} = 1,$$

which implies

$$C \neq 0 \quad \text{and} \quad F\{X : p(Z'C) = p(Z'0)\} = 1.$$

This contradicts to the condition (i) of (2.2).

Put $f_1(u) = 3\log p(u) + 3q(u)$, $f_2(u) = f_1(u) - p(u)q(u)$,

$$\begin{aligned} S_n^*(\gamma) &= \frac{1}{n}\sum_{i=1}^{n} f_1(Z_i'\gamma)Z_iZ_i', \\ S_n^{**}(\gamma) &= \frac{1}{n}\sum_{i=1}^{n} f_2(Z_i'\gamma)Z_iZ_i', \\ S^*(\gamma) &= \int f_1(Z'\gamma)ZZ'dF, \\ S^{**}(\gamma) &= \int f_2(Z'\gamma)ZZ'dF. \end{aligned} \tag{3.10}$$

It is easily seen that, the above four functions are all concave functions of γ. Under (ii) of (2.2), by SLLN,

$$\lim_{n\to\infty} S_n^*(\gamma) = S^*(\gamma) \qquad \text{a.s.} \tag{3.11}$$

$$\lim_{n\to\infty} S_n^{**}(\gamma) = S^{**}(\gamma) \qquad \text{a.s.}$$

Put

$$S_n(\underset{\sim}{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} p(\underset{\sim}{Z}_i'\underset{\sim}{\gamma})q(\underset{\sim}{Z}_i'\underset{\sim}{\gamma})Z_i Z_i' . \qquad (3.12)$$

We have

$$S_n(\underset{\sim}{\gamma}) = S_n^{*}(\underset{\sim}{\gamma}) - S_n^{**}(\underset{\sim}{\gamma}), \quad S(\underset{\sim}{\gamma}) = S^{*}(\underset{\sim}{\gamma}) - S^{**}(\underset{\sim}{\gamma}).$$

For any matrix $A = (a_{ij})_{0\leq i,j\leq p}$, write

$$||A|| = (\sum_{i,j=0}^{p} a_{ij}^2)^{1/2}.$$

Using (3.11) and Lemma 3.1, we get for all compact $D \subset R^{p+1}$,

$$\sup_{\underset{\sim}{\gamma}\in D} || S_n(\underset{\sim}{\gamma}) - S(\underset{\sim}{\gamma})|| \to 0 \quad \text{a.s.} \quad \text{as} \quad n \to \infty. \qquad (3.13)$$

Since $\hat{\underset{\sim}{\beta}}_n$ satisfies (3.6), we have

$$\frac{1}{n} \sum_{i=1}^{n} [Y_i - p(\underset{\sim}{Z}_i'\underset{\sim}{\beta})]Z_i = \frac{1}{n} \sum_{i=1}^{n} [p(\underset{\sim}{Z}_i'\hat{\underset{\sim}{\beta}}_n) - p(\underset{\sim}{Z}_i'\underset{\sim}{\beta})]Z_i$$

$$= \frac{1}{n} \sum_{i=1}^{n} p(\underset{\sim}{Z}_i'\underset{\sim}{\beta}_*)q(\underset{\sim}{Z}_i'\underset{\sim}{\beta}_*)Z_i Z_i'(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}) \qquad (3.14)$$

$$= S_n(\underset{\sim}{\beta}_*)(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}),$$

where $\underset{\sim}{\beta}_* = \lambda \underset{\sim}{\beta} + (1-\lambda)\hat{\underset{\sim}{\beta}}_n$ with some $\lambda \in (0,1)$. By (3.4) and (3.13),

$$S_n(\underset{\sim}{\beta}_*) - S(\underset{\sim}{\beta}) \to 0 \quad \text{a.s.} \quad \text{as } n \to \infty. \qquad (3.15)$$

From (3.8) and (3.15), it follows that with probability one for n large, $S_n(\underset{\sim}{\beta}_*) > 0$ and

$$S_n^{-1}(\underset{\sim}{\beta}_*) = S^{-1}(\underset{\sim}{\beta})\,(1+o(1)) \quad \text{a.s.} \qquad (3.16)$$

From (3.14) and (3.16), (3.7) follows. Lemma 3.4 is proved.

LEMMA 3.5. Define $H_n(\gamma)$ and $H(\ldots$ $\ldots .2)$. Under conditions (i) and (ii) of (2.2), we have

$$H_n(\hat{\beta}_n) - H_n(\beta) = -\frac{1}{2}(\hat{\beta}_n-\beta)'S(\beta)(\hat{\beta}_n-\beta) + o(\ldots^2) \quad \text{a.s.}$$

as $n \to \infty$, where $S(\gamma)$ is defined by (3.8).

Proof. By (3.2),

$$\frac{\partial H_n}{\partial \gamma'} = \frac{1}{n}\sum_{i=1}^{n}[p(Z_i'\beta)q(Z_i'\gamma) - q(Z_i'\beta)p(Z_i'\gamma)]Z_i',$$

$$\frac{\partial^2 H_n}{\partial\gamma\partial\gamma'} = -\frac{1}{n}\sum_{i=1}^{n} p(Z_i'\gamma)q(Z_i'\gamma)Z_iZ_i' = -S_n(\gamma),$$

which implies

$$\frac{\partial H_n}{\partial \gamma'}(\beta) = 0. \tag{3.18}$$

By the Taylor expansion,

$$\begin{aligned} H_n(\hat{\beta}_n) - H_n(\beta) &= \frac{\partial H_n}{\partial\gamma'}(\beta)(\hat{\beta}_n-\beta) + \frac{1}{2}(\hat{\beta}_n-\beta)'\frac{\partial^2 H_n}{\partial\gamma\partial\gamma'}(\beta^*)(\hat{\beta}_n-\beta) \\ &= -\frac{1}{2}(\hat{\beta}_n-\beta)'S_n(\beta^*)(\hat{\beta}_n-\beta), \end{aligned} \tag{3.19}$$

where $\beta^* = \tilde{\lambda}\beta + (1-\tilde{\lambda})\hat{\beta}_n$ for some $\tilde{\lambda} \in (0,1)$. Similar to (3.15), we have

$$\lim_{n\to\infty} S_n(\beta^*) = S(\beta) > 0 \quad \text{a.s.} \tag{3.15'}$$

as $n \to \infty$. The lemma follows from (3.19) and (3.15')

LEMMA 3.6. Under the conditions (i) and (ii) of (2.2), we have

$$\frac{1}{n}\log L_n(\hat{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n}[Y_i-p(Z_i'\beta)]Z_i'\beta + H_n(\beta)$$

$$+ \frac{1}{2}(\hat{\beta}_n - \beta)'S(\beta)(\hat{\beta}_n - \beta) + o(\|\hat{\beta}_n - \beta\|^2), \quad \text{a.s.} \tag{3.20}$$

as $n \to \infty$, where,

$$S(\beta) = \; p(Z'\beta)q(Z'\beta)ZZ'dF > 0. \tag{3.21}$$

and $H_n(\gamma)$ is defined in (3.2).

Proof. By (3.1), (3.2) and (2.1),

$$\frac{1}{n}\log L_n(\hat{\beta}_n) - H_n(\hat{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - p(Z_i'\beta))Z_i'\hat{\beta}_n$$

$$= \frac{1}{n}\sum_{i=1}^{n}[Y_i - p(Z_i'\beta)]Z_i'\beta$$

$$+ \frac{1}{n}\sum_{i=1}^{n}[Y_i - p(Z_i'\beta)]Z_i'(\hat{\beta}_n - \beta). \tag{3.22}$$

By (3,14), (3,15),

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i - p(Z_i'\beta)]Z_i'(\hat{\beta}_n - \beta)$$

$$= (\hat{\beta}_n - \beta)'S(\beta)(\hat{\beta}_n - \beta) + o(\|\hat{\beta}_n - \beta\|^2) \quad \text{a.s.} \tag{3.23}$$

as $n \to \infty$. (3.20) follows from Lemma 3.5 and (3.22), (3.23).

Now we take

$$B = \{0,2,3,\ldots,p\},$$

and put

$$\gamma_B' = (\gamma_0, \gamma_2, \gamma_3, \ldots, \gamma_p), \quad X_B' = (X^{(2)},\ldots,X^{(p)}), \quad X_{Bi}' = (X_i^{(2)},\ldots,X_i^{(p)}),$$

$$Z_B' = (1, X_B') = (1, X^{(2)},\ldots,X^{(p)}), \quad Z_{Bi}' = (1, X_{Bi}').$$

Write

$$\frac{1}{n}\log\tilde{L}_n(\underset{\sim}{\gamma}_B) = \frac{1}{n}\sum_{i=1}^{n}[Y_i\log p(\underset{\sim}{Z}_{Bi}'\underset{\sim}{\gamma}_B) + (1-Y_i)\log q(\underset{\sim}{Z}_{Bi}'\underset{\sim}{\gamma}_B)],$$

$$\tilde{H}(\underset{\sim}{\gamma}_B) = \int[p(\underset{\sim}{Z}'\underset{\sim}{\beta})\log p(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B) + q(\underset{\sim}{Z}'\underset{\sim}{\beta})\log q(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B)]dF, \tag{3.24}$$

where

$$p(u) = \frac{1}{1+e^{-u}}, \quad q(u) = 1-p(u).$$

Functions $\frac{1}{n}\log\tilde{L}_n(\underset{\sim}{\gamma}_B)$ and $\tilde{H}(\underset{\sim}{\gamma}_B)$ are all concave functions on R^p. Further

$$\frac{\partial\tilde{H}}{\partial\underset{\sim}{\gamma}_B} = \int[p(\underset{\sim}{Z}'\underset{\sim}{\beta})q(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B) - q(\underset{\sim}{Z}'\underset{\sim}{\beta})p(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B)]\underset{\sim}{Z}_B\, dF. \tag{3.25}$$

$$\frac{\partial^2\tilde{H}}{\partial\underset{\sim}{\gamma}_B\partial\underset{\sim}{\gamma}_B'} = -\int p(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B)q(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B)\underset{\sim}{Z}_B\underset{\sim}{Z}_B'\, dF. \tag{3.26}$$

Similar to the argument used in establishing (3.8), by (i) of (2.2) we have

$$\tilde{S}(\underset{\sim}{\gamma}_B) = -\frac{\partial^2\tilde{H}}{\partial\underset{\sim}{\gamma}_B\partial\underset{\sim}{\gamma}_B'} > 0, \tag{3.27}$$

Thus, $\tilde{H}(\underset{\sim}{\gamma}_B)$ is strictly concave. Since

$$\tilde{H}(\underset{\sim}{\gamma}_B) \leq \int[p(\underset{\sim}{Z}'\underset{\sim}{\beta})\log p(\underset{\sim}{Z}'\underset{\sim}{\beta}) + q(\underset{\sim}{Z}'\underset{\sim}{\beta})\log q(\underset{\sim}{Z}'\underset{\sim}{\beta})]dF < \infty, \tag{3.28}$$

$\tilde{H}(\underset{\sim}{\gamma}_B)$ has a unique maximum at some $\underset{\sim}{\gamma}_B^*$.

Assume that $\hat{\underset{\sim}{\gamma}}_B$ maximizes $\frac{1}{n}\log\tilde{L}_n(\underset{\sim}{\gamma}_B)$. By SLLN, for any $\underset{\sim}{\gamma}_B \in R^p$,

$$\lim_{n\to\infty}\frac{1}{n}\log\tilde{L}_n(\underset{\sim}{\gamma}_B) = \tilde{H}(\underset{\sim}{\gamma}_B) \qquad \text{a.s.} \tag{3.29}$$

By Lemmas 3.1 and 3.2, for any compact $D \subset R^p$,

$$\sup_{\underset{\sim}{\gamma}_B \in D}\left|\frac{1}{n}\log\tilde{L}_n(\underset{\sim}{\gamma}_B) - \tilde{H}(\underset{\sim}{\gamma}_B)\right| \to 0 \qquad \text{a.s.} \quad \text{as } n \to \infty, \tag{3.30}$$

and

$$\lim_{n\to\infty} \hat{\underset{\sim}{\gamma}}_B = \underset{\sim}{\gamma}_B^* \qquad \text{a.s.} \tag{3.31}$$

From (3.30) and (3.31), it follows that

$$\lim_{n\to\infty} \frac{1}{n} \log \tilde{L}_n(\hat{\underset{\sim}{\gamma}}_B) = \tilde{H}(\underset{\sim}{\gamma}_B^*) \qquad \text{a.s.} \tag{3.32}$$

Similar to the argument used in the beginning of the proof of Lemma 3.3, we get the following

LEMMA 3.7. Suppose that $\underset{\sim}{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ is the true parameter, $\beta_1 \neq 0$ and $B = \{0,2,3,\ldots,p\}$. Define $G_n(B)$ by (2.5). Then, under the conditions (i) and (ii) of (2.2), we have

$$\lim_{n\to\infty} \frac{1}{n} G_n(B) \overset{\text{a.s.}}{=} \tilde{H}(\underset{\sim}{\gamma}_B^*) < H(\underset{\sim}{\beta}). \tag{3.33}$$

4. THE PROOF OF THE THEOREMS

In the following, we only give the proof of the theorem 2.1. The proof of the theorem 2.2 is similar.

Assume that $\underset{\sim}{\beta}$ is the true parameter and B_0 is the best subset of A, $\{0\} \subset B_0$.

For any B A, $B = \{j_0, j_1, \ldots, j_s\}$, where $j_0 = 0 < j_1 < \ldots < j_s$; put

$$\underset{\sim}{X}_B' = (X^{(j_1)}, \ldots, X^{(j_s)}), \quad \underset{\sim}{X}_{Bi}' = (X_i^{(j_1)}, \ldots, X_i^{(j_s)})$$

$$\underset{\sim}{\beta}_B' = (\beta_0, \beta_{j_1}, \ldots, \beta_{j_s}), \quad \underset{\sim}{\gamma}_B' = (\gamma_0, \gamma_{j_1}, \ldots, \gamma_{j_s}), \tag{4.1}$$

$$\underset{\sim}{Z}_B' = (1, \underset{\sim}{X}_B'), \quad \underset{\sim}{Z}_{Bi}' = (1, \underset{\sim}{X}_{Bi}'),$$

and denote by F_B the distribution of $\underset{\sim}{X}_B$. It is easily seen that, if $\underset{\sim}{\beta}_B \neq \underset{\sim}{\gamma}_B$, then

$$F_B\{\underset{\sim}{X}_B : p(\underset{\sim}{Z}_B'\underset{\sim}{\beta}_B) \neq p(\underset{\sim}{Z}_B'\underset{\sim}{\gamma}_B)\} > 0. \tag{4.2}$$

Now assume that $B_0 \subset B \subset A$, $B \neq B_0$, then $\#(B) > \#(B_0)$. By (4.2) and (ii) of (2.2), using Lemmas 3.4 and 3.6, we have

$$G_n(B_0) = n\, W_{n,B_0}(\underset{\sim}{\beta}_{B_0}) + n\, H_{n,B_0}(\underset{\sim}{\beta}_{B_0}) + O(\log\log n) \quad \text{a.s.},$$
$$\tag{4.3}$$
$$G_n(B) = n\, W_{n,B}(\underset{\sim}{\beta}_B) + n\, H_{n,B}(\underset{\sim}{\beta}_B) + O(\log\log n) \quad \text{a.s.}$$

where

$$W_{n,B}(\underset{\sim}{\beta}_B) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - p(\underset{\sim}{Z}_{Bi}'\underset{\sim}{\beta}_B)]\underset{\sim}{Z}_{Bi}'\underset{\sim}{\beta}_B,$$

$$H_{n,B}(\underset{\sim}{\beta}_B) = \frac{1}{n}\sum_{i=1}^{n}[p(\underset{\sim}{Z}'_{Bi}\underset{\sim}{\beta}_B)\log p(\underset{\sim}{Z}'_{Bi}\underset{\sim}{\beta}_B) + \tag{4.4}$$

$$+ q(\underset{\sim}{Z}'_{Bi}\underset{\sim}{\beta}_B)\log q(\underset{\sim}{Z}'_{Bi}\underset{\sim}{\beta}_B)].$$

Since $\beta_i = 0$ for $i \bar{\in} B_0$, we have

$$W_{n,B}(\underset{\sim}{\beta}_B) = W_{n,B_0}(\underset{\sim}{\beta}_{B_0}), \quad H_{n,B}(\underset{\sim}{\beta}_B) = H_{n,B_0}(\underset{\sim}{\beta}_{B_0}). \tag{4.5}$$

By (2.6), (2.7), (4.3) and (4.5), with probability one for n large,

$$I_n(B_0) - I_n(B) \geq G_n(B_0) - G_n(B) + C_n$$

$$= O(\log\log n) + C_n > 0. \tag{4.6}$$

Further, using (4.5) and Lemma 3.3, we have

$$\lim_{n\to\infty}\frac{1}{n}G_n(B_0) = \lim_{n\to\infty}\frac{1}{n}G_n(A) = H(\underset{\sim}{\beta}) \quad \text{a.s.} \tag{4.7}$$

Now we assume that $\{0\} \subseteq B \subseteq A$ and there exists some integer i such that $i \in B_0$ and $i \bar{\in} B$. Without loss of generality, we can assume that $i = 1$.

Put

$$B_1 = \{0,2,3,\ldots,p\}.$$

By Lemma 3.7, we have

$$\limsup_{n\to\infty}\frac{1}{n}G_n(B) \leq \lim_{n\to\infty}\frac{1}{n}G_n(B_1)$$

$$= \tilde{H}(\underset{\sim}{\gamma}^*_{B_1}) < H(\underset{\sim}{\beta}) \quad \text{a.s.} \tag{4.8}$$

where $\underset{\sim}{\gamma}^*_{B_1}$ maximizes $\tilde{H}(\underset{\sim}{\gamma}_{B_1})$. By (4.7), (4.8) and $\lim_{n\to\infty} C_n/n = 0$, with probability for n large,

$$I_n(B_0) - I_n(B) \geq G_n(B_0) - G_n(B_1) + O(C_n)$$

$$\geq \frac{n}{2}(H(\underset{\sim}{\beta}) - \tilde{H}(\underset{\sim}{\gamma}^*_{B_1})) + O(C_n) > 0. \tag{4.9}$$

From (4.6) and (4.9), it follows that, with probability one for n large

$$\hat{B} = B_0.$$

That is the desired.

REFERENCES

[1] ANDERSON, J A. (1982). Logistic discrimination. *Handbook of Statistics* (P R Krishnaiah and L. N. Kanal, eds.), North-Holland Publishing Company, Vol 2, 169-191.

[2] ANDERSON, P K and Gill, R D (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, 10, 1100-1120.

[3] BERKSON, J. (1951). Why I prefer logits to probits. *Biometrics* 7, 327-339.

[4] COX, D R (1970). *Analysis of Binary Data*. Chapman and Hall, London.

[5] EFRON, B (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.*, 70, 892-898.

[6] GORDON, T. and KANNEL, W. E. (1968). *Introduction and General Background in the Framingham Study - The Framingham Study, Sections 1 and 2*. National Heart, Lung, and Blood Institute, Bethesda, Maryland.

[7] PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist.*, 9, 705-724.

[8] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

[9] STEFANSKI, L. A. (1983). Influence and measurement error in logistic regression. *Institute of Statistics Mimeo Series*, No. 1548, University of North Carolina, Chapel Hill.

END
DATE
FILMED
DEC.
1987