Ann. Inst. Statist. Math.
36 (1984), Part A, 363-373

THE MAXIMUM FULL AND PARTIAL LIKELIHOOD ESTIMATORS IN THE PROPORTIONAL HAZARD MODEL

TAKEMI YANAGIMOTO AND TOSHINARI KAMAKURA

(Received Sept. 6, 1983; revised Nov. 9, 1983)

Summary

The maximum full likelihood estimator in the proportional hazard model is explored in relation to the maximum partial likelihood estimator. In the scalar parameter case both the estimators have a common sign, and the absolute value of the former is strictly greater than that of the latter except for trivial cases. We point out also that the maximum full likelihood estimator after a simple modification of the likelihood equation provides a good approximation to the maximum partial likelihood estimator. Similar results are valid for the likelihood ratio tests.

1. Introduction

The proportional hazard model is widely accepted in survival analysis and the partial likelihood is employed to estimate the structure parameter (Cox [7], [8]). We know (for example Miller [20], pp. 57–59) that the (generalized) full likelihood is applicable to estimate the structure parameter as well as nuisance parameters. Less attention has been given to the full likelihood.

Analytically the same problems appear in various fields of statistics. Prentice-Breslow [23] and Farwell [9] remarked that the inference procedure using the logistic model contains the same problems in case-control studies where data are summarized in multiple 2×2 or $k\times2$ tables. The proportional hazard model provides a type of logistic model for the contingency table with ordered categories (Pregibon [22]). As an extension of the proportional hazard model, the proportional intensity model in the point process is employed to describe an asthma attack in relation to environmental factors (Korn-Whittmore [17], Yanagimoto-Kamakura [25], published in Japanese). Though the partial likelihood becomes the conditional likelihood, we will use the term of partial likelihood even in such cases for convenience.

Partial and full likelihoods are basic criteria for deriving estimators and test statistics. Simulation results conducted by Mckinlay [19], Lubin [18] and Hauck et al. [14] concluded that the maximum partial likelihood estimator is superior to the maximum full likelihood estimator. Andersen [2] gave a sufficient condition under which the maximum partial likelihood estimator is consistent. The maximum full likelihood procedure, however, has some advantages. It provides estimators of all parameters in a model and the maximized likelihood, which permits us to compare the goodness of fit of the proportional hazard model with those of other candidate models (see Nelder-Wedderburn [21], Akaike [1]). Also we can encounter computation difficulties in obtaining the maximum partial likelihood estimator.

It is worthwhile to explore the behavior of the maximum full likelihood estimator, even when the maximum partial likelihood estimator is applicable. Obviously both estimators behave similarly in a rough sense, but they are different in detail. Findings on differences between the two estimators should be helpful in choosing one of the two. In this paper we prove that the maximum full and partial likelihood estimators for any sample have a common sign and that the absolute value of the former is strictly greater than that of the latter except for trivial cases. In addition the maximum partial likelihood estimator is fairly well approximated by the modified maximum full likelihood estimator, where the proposed modification depends simply on the numbers of individuals at risk.

The parallel inequality and approximation are valid for the full and partial likelihood ratio test statistics. Throughout the paper we focus our attention on a scalar parameter case, unless otherwise specified.

2. Partial likelihood and full likelihood

The partial likelihood is written by the product, over each distinct failure time, of terms

$$P(\beta)=\prod_{i=1}^{r} \exp(\beta z_i)\Big/\sum_{\Psi}\prod_{i=1}^{r} \exp(\beta z_{\psi_i}),$$

where $z_1,\cdots,z_n$ denote covariate vectors for the n individuals at risk at that failure time, $z_1,\cdots,z_r$ correspond to the failures and Ψ denotes the set of all subsets $\psi=\{\psi_1,\cdots,\psi_r\}$ of size r from $\{1,\cdots,n\}$. Throughout the paper we assume that all of $z_1,\cdots,z_n$ are not identical and r is an integer between 1 and $n-1$. To evaluate the maximum partial likelihood estimator we explore basic properties and characteristics of $P(\beta)$. The following transformation of $P(\beta)$ is convenient for the calculation;

$$\log P(\beta)=\sum_{i=1}^{r} \beta z_i-\log (\sum_{\Psi} \exp \beta \sum_{\phi} z_j) .$$

The maximum partial likelihood estimator $\hat{\beta}_p$ is a root of the equation

$$LP(\beta)=\sum_{i=1}^{r} z_i-(\sum_{\Psi} \sum_{\phi} z_j \exp \beta \sum_{\phi} z_j)/\sum_{\Psi} \exp \beta \sum_{\phi} z_j=0 .$$

The (generalized) full likelihood is expressed by the product, over each distinct failure time, of terms

$$L(\beta, \lambda)=\prod_{i=1}^{r} \lambda \exp \beta z_i \Big/ \prod_{i=1}^{n} (1+\lambda \exp \beta z_i) .$$

Differentiating $\log L(\beta, \lambda)$ with respect to β and λ to obtain the maximum full likelihood estimator, we get

$$\sum_{i=1}^{r} z_i-\sum_{i=1}^{n} \lambda z_i \exp \beta z_i/(1+\lambda \exp \beta z_i)=0$$

and

$$r/\lambda-\sum_{i=1}^{n} \exp \beta z_i/(1+\lambda \exp \beta z_i)=0 .$$

From the latter equation $\lambda(\beta)$ is uniquely determined for any fixed β. Using $\lambda(\beta)$, we define

$$LF(\beta)=\sum_{i=1}^{r} z_i-\sum_{i=1}^{n} \lambda(\beta) z_i \exp \beta z_i/\{1+\lambda(\beta) \exp \beta z_i\} .$$

The maximum full likelihood estimator, $\hat{\beta}_F$, is a root of the equation $LF(\beta)=0$. We will denote $\lambda(\beta)$ by λ for simplicity.

We remark here that the whole likelihoods are the products over all distinct failure times T. Thus the likelihood equations in the strict sense are $\sum LP_t(\beta)=0$ and $\sum LF_t(\beta)=0$, where the summations extend over t in T. As far as we are concerned, the results in a single failure time can extend straightforwardly to those in multiple failure times. We will focus on the likelihood equations of a single failure time and suppress the suffix t, unless any confusion occurs.

The comparison of behaviors $LF(\beta)$ and $LP(\beta)$ provides us with those of the maximum full and partial likelihood estimators. Though both the functions behave similarly, they have different behaviors in detail. The differences may give rise to sharp differences between the two estimators.

PROPOSITION 1. Let $K(\beta)$ be either of $LF(\beta)$ or $LP(\beta)$. Denote $\sum_{i=1}^{n} z_i/n$ by $\bar{z}$, and $z_{(1)}+\cdots+z_{(r)}$ and $z_{(n-r+1)}+\cdots+z_{(n)}$ by $L(z; r)$ and $U(z; r)$ respectively, where $z_{(1)}, \cdots, z_{(n)}$ are ordered covariates in the ascending order. Then $K(\beta)$ has the following four properties:

(i) $K(0)=z_1+\cdots+z_r-r\bar{z}$.
(ii) $K'(\beta)$ is negative for any β, that is, $K(\beta)$ is strictly decreasing.
(iii) $\lim_{\beta\to-\infty} K(\beta)=U(z;r)$.
(iv) $\lim_{\beta\to\infty} K(\beta)=L(z;r)$.

The extension to the case of the vector parameter β is straightforward. From Proposition 1 it follows that if either of the two estimators exists, then the other also exists, and they are uniquely determined. Furthermore both the estimators have a common sign.

3. Inequality between estimators

Simulation studies by Farewell-Prentice [10], Lubin [18], Hauck et al. [14] and Mckinlay [19] in the situation of a case-control study conclude that $\hat{\beta}_P$ behaves fairly well while $\hat{\beta}_F$ tends to have upward bias. Kamakura-Yanagimoto [16] conjectured, based on their numerical experiments, that $|\hat{\beta}_F|$ is strictly greater than $|\hat{\beta}_P|$. The following theorem provides a theoretical basis of these previous observations. The proof will be presented in Appendix.

THEOREM. *Suppose that* $\sum(z_i-\bar{z})^2\neq 0$. *Then the functions* $LP(\beta)$ *and* $LF(\beta)$ *have a unique intersection at* $\beta=0$. *It holds also that* $LP(\beta)<LF(\beta)$ *for* $\beta>0$. *The reverse inequality is valid for* $\beta<0$.

Remark 1. The above theorem proves affirmatively the conjecture in Kamakura-Yanagimoto [16]. In fact Proposition 1 shows that if either of $\hat{\beta}_F$ and $\hat{\beta}_P$ exists, the other also exists. Suppose $LP(0)=LF(0)>0$. Then it follows that $\hat{\beta}_F>\hat{\beta}_P>0$. The reverse inequalities hold, when $LP(0)<0$.

They give inequalities between $LP(\beta)$ and $LB(\beta)$, where $LB(\beta)$ is

$$LB(\beta)=\sum_{i=1}^{r} z_i-r\left(\sum_{i=1}^{n} z_i \exp \beta z_i \Big/ \sum_{i=1}^{n} \exp \beta z_i\right) .$$

In fact it holds that $\hat{\beta}_P\geqq\hat{\beta}_B>0$ when $LB(0)>0$, where $LB(\hat{\beta}_B)=0$. An alternative proof of these inequalities is available along the line of the proof in Appendix.

COROLLARY. *Suppose the set of covariates consists of* k 1's *and* $(n-k)$ 0's. *Let* λ *is the solution of* $k\lambda e^{\beta}/(1+\lambda e^{\beta})+(n-k)\lambda/(1+\lambda)=r$. *Then it follows that for* $\beta<0$

$$LP(\beta)-LF(\beta)=k\frac{\lambda e^{\beta}}{1+\lambda e^{\beta}}-\frac{\sum\binom{k}{u}\binom{n-k}{r-u}ue^{u\beta}}{\sum\binom{k}{u}\binom{n-k}{r-u}e^{u\beta}}>0 .$$

This is the former part of (7) *in Harkness* [13], *where he gave the inequality as a bound of the mean of the extended hypergeometric distribution.*

Next consider the test problem for $\beta=0$ against $\beta\neq 0$. The commonly used test statistics would be the full and partial likelihood ratio test statistics, χ_F^2 and χ_P^2, which are expressed by

$$\chi_F^2=-2\int_0^{\hat\beta_F} LF(\beta)d\beta$$

and

$$\chi_P^2=-2\int_0^{\hat\beta_P} LP(\beta)d\beta ,$$

respectively. The usual manner to define the rejection region is χ_F^2 $(\chi_P^2)>\chi_\alpha^2$, where χ_α^2 denotes the upper αth point for a specified level α $(1>\alpha>0)$. Obviously the above theorem implies that $P(\chi_F^2>\chi_\alpha^2)>P(\chi_P^2>\chi_\alpha^2)$ for any α. Following the proof of the corollary in Kamakura-Yanagimoto [16], we can extend this inequality to the multidimensional case of β.

4. Expansion near the origin

To compare the behaviors of $LF(\beta)$ and $LP(\beta)$ quantitatively, we present in this section their power expansions near the origin. Since both the functions behave similarly as shown in the previous sections, we expect the quantitative difference near the origin may be critical in a wide range of β. The behavior near the origin itself is of practical importance for studying the estimator and test procedure.

PROPOSITION 2. The power expansions of $LF(\beta)$ and $LP(\beta)$ near the origin up to the third order are as follows: for $n\geqq 4$

(i) $$LF(\beta)\fallingdotseq\sum_{i=1}^{r} z_i-\Big[r\bar z+\frac{r(n-r)}{n^2}s_2\beta+\frac{1}{2}\frac{r(n-r)(n-2r)}{n^3}s_3\beta^2+\frac{1}{6}\frac{r(n-r)}{n^5}\{n(n^2-6rn+6r^2)s_4-3(n-2r)^2s_2^2\}\beta^3\Big] ,$$

(ii) (Cox [6], p. 62)

$$LP(\beta)\fallingdotseq\sum_{i=1}^{r} z_i-\Big[r\bar z+\frac{r(n-r)}{n(n-1)}s_2\beta+\frac{1}{2}\frac{r(n-r)(n-2r)}{n(n-1)(n-2)}s_3\beta^2$$

$$+\frac{1}{6}\frac{r(n-r)}{n^2(n-1)(n-2)(n-3)}\{n(n^2-6rn+6r^2+n)s_4$$

$$\left.+3(r-1)n(n-r-1)s_2^2\}\beta^3\right],$$

where $s_k=\sum(z_i-\bar{z})^k$, $k=2,3$ and 4.

The function $LF(\beta)$ has steeper slope near the origin than $LP(\beta)$. The relative ratio is $n/(n-1)$, which means that $LF(n\beta/(n-1))$ is close to $LP(\beta)$ near the origin. The power expansion of $LA(\beta)=LF(n\beta/(n-1))$ is expressed by

$$LA(\beta)\doteqdot\sum_{i=1}^{r}z_i-\left\{r\bar{z}+\frac{r(n-r)}{n(n-1)}s_2\beta+\left(\frac{n}{n-1}\right)^2c_3\beta^2+\left(\frac{n}{n-1}\right)^3c_4\beta^3\right\},$$

where c_3 and c_4 are coefficients of order 2 and 3 of $LF(\beta)$. Though $LA(\beta)$ is defined so as to adjust the coefficient of $LF(\beta)$ of order 1 to that of $LP(\beta)$, the coefficient of order 2 of $LA(\beta)$ becomes to be closer to that of $LP(\beta)$ than that of $LF(\beta)$.

When the absolute values of $\hat{\beta}_F$ and $\hat{\beta}_P$ are small, then $\hat{\beta}_P$ can be approximated by $(n-1)\hat{\beta}_F/n$. Under this condition the relative difference between the estimators is estimated as $1/n$. From the same reasoning we obtain $\chi_P^2\doteqdot(n-1)\chi_F^2/n$ when the absolute value of $\hat{\beta}_F$ is small.

5. Diagnosis of the approximation

We discuss in this section the accuracy of the approximation of $LA(\beta)$ and $\hat{\beta}_A=(n-1)\hat{\beta}_F/n$ to $LP(\beta)$ and $\hat{\beta}_P$, respectively. The extension to the case of multiple failure times is straightforward. In fact we may define $LA(\beta)=\sum LA_t(\beta)$ by $LA(\beta)=\sum LF_t\{(n_t-1)\beta/n_t\}$, where n_t stands for the number of individuals at risk at a failure time t. Under the condition that all n_t's are common, $LA(\beta)$ is written as $LF((n-1)\beta/n)$. The above condition is often satisfied in the case of case-control studies and the point process describing asthma attacks.

Breslow [4] obtained a similar result. Since his interest focused on the analysis of multiple 2×2 tables, he assumed that covariates are binary. Under the condition that $\lambda_0/(1-\lambda_0)=0.3$, he compared $\mathrm{E}((n-1)\hat{\beta}_F/n\,|\,\lambda_0,\beta_0)$ and $\mathrm{E}(\hat{\beta}_P\,|\,\lambda_0,\beta_0)$ for various β_0. His conclusion is that they are quite close to each other, especially when $|\beta_0|<\log 2$. The result gives an insight into the relationship when the number of 2×2 tables (number of failure times) is large. We assert that his result is also valid for any sample.

We can check the accuracy of the approximation using the simulation results by Lubin [18] and Farewell-Prentice [10]. (They contain typographical errors; personal communications with the authors). They assumed that the numbers of individuals at risk, which are the numbers of patients in each stratum, in their paper, are common. We find the above approximation is fairly good. The proposed approximated estimator is in general more accurate than those proposed by Farewell-Prentice [10] and Breslow [3].

The above findings on the accuracy of the approximation are available for χ_P^2 and $\chi_A^2=2\int LA(\beta)d\beta$, where the integral ranges from 0 to $\hat{\beta}_A$. Usually in the test procedure our interest centers on the moderately small χ_A^2 or χ_F^2, since the significance of the test is often the main concern.

The proposed approximated estimator and test statistic should be helpful in the cases of multiple 2×2 tables and contingency table when both numbers, n and r, are large. There is generally no serious problem in obtaining $LA(\beta)$. However we may encounter computational difficulties in obtaining $LP(\beta)$, such as a need for too much CPU time as well as frequent overflow errors. The difficulties are serious, especially when access to computing facilities is restricted.

The following example illustrates the above approximation.

Table 1. Hypothetical data, $t_1<t_2<t_3$.

	Covariate	No. of individuals		
		t_1	t_2	t_3
Group 1	0	$2K$	$2K$	$4K$
Group 2	1	$2K$	$3K$	$2K$
Group 3	2	$3K$	$2K$	K

Example. Consider the hypothetical data in Table 1. We obtain the estimators of β for a covariate (0, 1, 2) and the chi-square test statistics for $\beta=0$ against $\beta\neq 0$ based on $LP(\beta)$, $LA(\beta)$ and $LF(\beta)$. Following Gail et al. [11] the algorithm for calculating $LP(\beta)$ and its derivatives is coded. The results for K ranging from 1 to 10, along with the CPU times are presented in Table 2.

As expected, the approximation of $\hat{\beta}_A$ and χ_A^2 to $\hat{\beta}_P$ and χ_P^2 is satisfactory. The absolute value of the difference between the estimated values decreases from .00112 to .00012 as K increases, and that between the chi-square values remains close to .002. We observe a relatively large gap between $\hat{\beta}_P$ and $\hat{\beta}_F$ even for $K=10$. The CPU time for obtaining $\hat{\beta}_P$ increases with K, while that for $\hat{\beta}_F$ is about .007 second for any K, using a HITAC M-200H machine.

Table 2. Comparisons of the maximum partial, full and adjusted likelihood estimates and their likelihood ratio test statistics with CPU times.

K	$\hat{\beta}_P$	χ^2_P	(CPU time)	$\hat{\beta}_A$	χ^2_A	(CPU time)	χ^2_F
1	.59317	1.85227	(.012)	.59429	1.85415	(.007)	1.96687
2	.61178	3.81921	(.037)	.61236	3.82115	(.007)	3.93374
3	.61798	5.78609	(.079)	.61838	5.78806	(.007)	5.90061
4	.62108	7.75296	(.136)	.62138	7.75494	(.007)	7.86748
5	.62294	9.71983	(.211)	.62318	9.72182	(.007)	9.83435
6	.62418	11.68669	(.302)	.62439	11.68869	(.007)	11.80122
7	.62507	13.65356	(.409)	.62524	13.65556	(.007)	13.76809
8	.62574	15.62043	(.533)	.62589	15.62243	(.007)	15.73496
9	.62625	17.58730	(.670)	.62639	17.58930	(.007)	17.70183
10	.62667	19.55416	(.825)	.62679	19.55616	(.007)	19.66870
$\hat{\beta}_F$	.63039			.63039			

Appendix: Proof of Theorem

To use some properties of symmetric functions, we introduce the following notation:

(i) $y_i=u^{z_i}=\exp\beta z_i$.

(ii) $C_r=\sum_{\phi\in\Psi_r}\exp\beta\sum_{j\in\phi}z_j\left(=\sum_{\phi\in\Psi_r}u^{\sum_{j\in\phi}z_j}=\sum_{\phi\in\Psi_r}\prod_{j\in\phi}y_j\right)$.

(iii) $C_r(i)=\sum_{\phi\in\Psi_r(i)}\exp\beta\sum_{j\in\phi}z_j$ for $i=1,\cdots,n$.

(iv) $C_r(i,j)=\sum_{\phi\in\Psi_r(i,j)}\exp\beta\sum_{k\in\phi}z_k$ for $i,j=1,\cdots,n$, $i\neq j$,

where Ψ_r denotes the set of all combinations of size r from the set $\{1,\cdots,n\}$, and $\Psi_r(i)$ and $\Psi_r(i,j)$ are those from $\{1,\cdots,n\}$-$\{i\}$ and $\{1,\cdots,n\}$-$\{i,j\}$, respectively.

Remark that

$$\text{(A1)}\qquad \prod_{i=1}^{n}(x+y_i)=\sum_{r=0}^{n}C_r x^{n-r}$$

and

$$\text{(A2)}\qquad \left\{\prod_{j=1}^{n}(x+y_j)\right\}\Big/(x+y_i)=\sum_{r=0}^{n-1}C_r(i)x^{n-r-1}\,.$$

We prepare the following three lemmas.

LEMMA 1 (*Theorem* 53 *in Hardy et al.* [12]). *For* $n-1\geqq r\geqq 1$ *it holds that* $C_r^2>C_{r+1}C_{r-1}$.

The following definitions in special cases permit us to delete the inessential exceptions in the subsequent identities. We define $C_0=C_0(i)=1$. Also we define for $C_{-1}(i,j)=0$ and $C_0(i,j)=1$ for $n>3$ and any i

and j, and $C_s(i, j)=0$ for $n=2$ and any s, i and j.

LEMMA 2. (i) (Howard [15]) $C_r=y_iC_{r-1}(i)+C_r(i)$.

(ii) $rC_r=\sum_{i=1}^{n} y_iC_{r-1}(i)$.

(iii) $\sum_{\phi\in\Psi_r}\sum_{j\in\phi} z_j u^{\sum_{j\in\phi} z_j}=\sum_{i=1}^{n} z_iu^{z_i}C_{r-1}(i)$.

PROOF. First we note that (A1) is rewritten by

$$\prod_{i=1}^{n}(x+u^{z_i})=\sum_{r=0}^{n} C_rx^{n-r}\,. \tag{A3}$$

Differentiating both sides of (A1) by x and applying (A2), we obtain $(n-r)C_r=\sum C_r(i)$ by comparing coefficients of order $n-r-1$. This together with (i) derives (ii). Next to show (iii) we remark here that

$$\frac{d}{du}C_r=\sum_{\phi\in\Psi_r}\sum_{j\in\phi} z_ju^{(\sum_{j\in\phi} z_j)-1}\,.$$

Differentiating both the sides of (A3) by u and applying (A2), we obtain (iii) by comparing coefficients of order $n-r$.

LEMMA 3. *Let* $\{p_i\}$ *and* $\{q_i\}$ *be two probability functions on* $\{z_i\}$. *Suppose that* $z_1\leqq\cdots\leqq z_n$ *and that* $1/r>p_i$, $q_i>0$ *for* $i=1,\cdots,n$. *If it holds that*

$$g(i)=\{rp_i/(1-rp_i)\}/\{rq_i/(1-rq_i)\} \tag{A4}$$

is nondecreasing in i, *then* $\sum z_ip_i\geqq\sum z_iq_i$, *where the equality holds only when* $\sum(z_i-\bar{z})^2=0$ *or* $\sum(p_i-q_i)^2=0$.

PROOF. Since $rx/(1-rx)$ is strictly increasing in x for $1/r>x>0$, it holds that $\operatorname{sgn}(g(i)-1)=\operatorname{sgn}(p_i-q_i)$, where $\operatorname{sgn}(x)=1$ for $x>0$, $=0$ for $x=0$, and $=-1$ for $x<0$. Let i_0 be the minimum integer i such that $g(i)-1\geqq 0$. Then it follows that $p_i\geqq q_i$ for $i\geqq i_0$ and $p_i<q_i$ for $i<i_0$, which implies that

$$\sum_{j=i}^{n} p_j\geqq\sum_{j=i}^{n} q_j \qquad \text{for any } i\,. \tag{A5}$$

Assume that $\sum(p_i-q_i)^2\neq 0$. Then it holds that $p_n>q_n$ and $p_1<q_1$, which imply the strict inequality in (A5). Applying the identity $\sum z_ip_i-\sum z_iq_i=\sum_{i=2}^{n}\left\{(z_i-z_{i-1})\sum_{j=i}^{n}(p_j-q_j)\right\}$, we complete the proof.

Next we move to the proof of the Theorem. We will show only the case for $\beta>0$ and $z_1\leqq\cdots\leqq z_n$. Parallel discussions derive the proof in other cases. Set

$$p_i = y_i C_{r-1}(i)/rC_r$$

and

$$q_i = \lambda y_i / r(1+\lambda y_i) .$$

Then the inequality to be proved is rewritten by $\sum z_i p_i > \sum z_i q_i$. Lemma 2 (ii) and the definition of λ derive $\sum p_i = 1$ and $\sum q_i = 1$, respectively. It is easily checked that $1/r > p_i$, $q_i > 0$ for any i. Define $g(i)$ by (A4) in Lemma 3. Then Lemma 2 (i) implies that $g(i)$ is broken down to

$$g(i) = \{y_i C_{r-1}(i)/(C_r - y_i C_{r-1}(i)\} / \lambda y_i = C_{r-1}(i)/C_r(i)\lambda .$$

Applying again Lemma 1 (i), it follows that

$$\begin{aligned} g(i)-g(j) &= \{C_{r-1}(i)C_r(j) - C_r(i)C_{r-1}(j)\} / \lambda C_r(i)C_r(j) \\ &= [\{y_j C_{r-2}(i, j) + C_{r-1}(i, j)\}\{y_i C_{r-1}(i, j) + C_r(i, j)\} \\ &\quad - \{y_j C_{r-1}(i, j) + C_r(i, j)\}\{y_i C_{r-2}(i, j) + C_{r-1}(i, j)\}] / \lambda C_r(i)C_r(j) \\ &= (y_i - y_j)\{C_{r-1}^2(i, j) - C_r(i, j)C_{r-2}(i, j)\} / \lambda C_r(i)C_r(j) . \end{aligned}$$

Remark that Lemma 1 implies $\mathrm{sgn}\,(g(i)-g(j)) = \mathrm{sgn}\,(y_i - y_j)$. This together with Lemma 3 completes the proof.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd Inter. Symp. on Information Theory* (eds. Petrov, B. N. and Csaki, F.), Akademia Kiado, Budapest, 267-281.

[2] Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators, *J. R. Statist. Soc.*, B, **32**, 283-301.

[3] Breslow, N. (1972). Discussion on Professor Cox's paper, *J. R. Statist. Soc.*, B, **34**, 187-220.

[4] Breslow, N. (1981). Odds ratio estimators when the data are sparse, *Biometrika*, **68**, 73-84.

[5] Breslow, N. (1982). Covariance adjustment of relative-risk estimates in matched studies, *Biometrics*, **38**, 661-672.

[6] Cox, D. R. (1970). *The Analysis of Binary Data*, Methuen & Co., London.

[7] Cox, D. R. (1972). Regression models and life tables (with discussion), *J. R. Statist. Soc.*, B, **34**, 187-220.

[8] Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269-276.

[9] Farewell, V. T. (1979). Some results on the estimation of logistic models based on retrospective data, *Biometrika*, **66**, 27-32.

[10] Farewell, V. T. and Prentice, R. L. (1980). The approximation of partial likelihood with emphasis on case-control studies, *Biometrika*, **67**, 273-278.

[11] Gail, M. H., Lubin, J. H. and Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times, *Biometrika*, **68**, 703-707.

[12] Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities*, University Press, Cambridge.

[13] Harkness, W. L. (1965). Properties of the extended hypergeometric distribution, *Ann. Math. Statist.*, **36**, 938-945.

[14] Hauck, W. W., Anderson, S. and Leahy, F. L. (1982). Finite-sample properties of some old and some new estimators of a common odds ratio from multiple 2×2 tables, *J. Amer. Statist. Ass.*, **77**, 145-152.

[15] Howard, S. (1972). Discussion on Professor Cox's paper, *J. R. Statist. Soc.*, B, **34**, 187-220.

[16] Kamakura, T. and Yanagimoto, T. (1983). Evaluation of the regression parameter estimators in the proportional hazard model, *Biometrika*, **70**, 530-533.

[17] Korn, E. L. and Whittemoore, A. S. (1979). Methods for analysing panel studies of acute health effects of air pollution, *Biometrics*, **35**, 795-802.

[18] Lubin, J. H. (1981). An empirical evaluation of the use of conditional and unconditional likelihoods for case-control data, *Biometrika*, **68**, 567-571.

[19] Mckinlay, S. M. (1975). The effect of bias on estimators of relative risk for pair-matched and stratiffied samples, *J. Amer. Statist. Ass.*, **70**, 859-864.

[20] Miller, R. G. (1980). *Survival Analysis*, Wiley, New York.

[21] Nelder, W. and Wedderburn, R. W. M (1972). Generalized linear model, *J. R. Statist. Soc.*, A, **135**, 370-384.

[22] Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications, *Biometrics*, **38**, 485-498.

[23] Prentice, R. L. and Breslow, N. E. (1978). Retrospective studies and failure time models, *Biometrika*, **65**, 153-158.

[24] Thomas, M. A. and Taub, A. E. (1982). Calculating binomial probabilities when the trial probabilities are unequal, *Statist. Comput. & Simul.*, **14**, 125-131.

[25] Yanagimoto, T. and Kamakura, T. (1983). Analysis for the longitudinal data using point process models—Case of the evaluation of environmental factors to the asthma attacks (in Japanese), *Jap. J. Appl. Statist.*, **11**, 39-47.