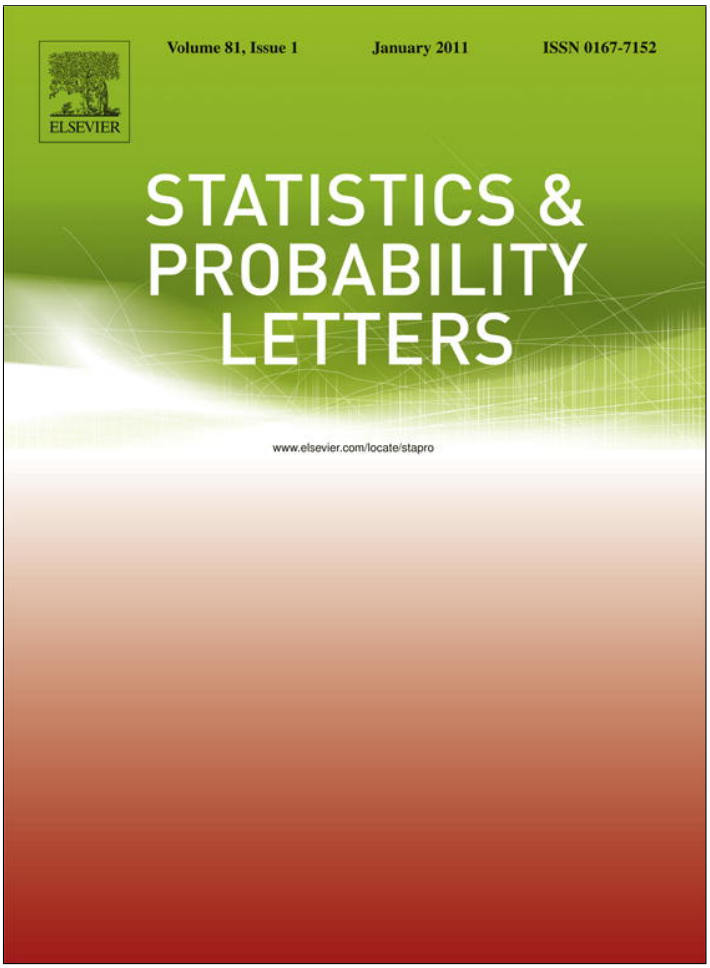

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Statistics and Probability Letters 81 (2011) 1802–1807

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Nonparametric estimation of the log odds ratio for sparse data by kernel smoothing

Ziqi Chen *, Ning-Zhong Shi, Wei Gao

Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, 130024, China

ARTICLE INFO

Article history:
Received 30 November 2010
Received in revised form 19 June 2011
Accepted 26 June 2011
Available online 6 July 2011

Keywords:
Mantel–Haenszel estimating function
Odds ratio
Sparse data

ABSTRACT

Regression analysis of the odds ratios for sparse data has received a lot of attention. However, existing works are restricted to the parametric case, and a parametric model may be a misspecification, which may lead to biased and inefficient estimators. Little attention is received for nonparametric regression analysis of the odds ratios. Based on kernel smoothing techniques, we propose two simple estimators of the log odds-ratio function for sparse data. Large sample properties of the estimators are derived, and the methods proposed are evaluated through simulation.

1. Introduction

Odds ratio is the key parameter for biomedical studies of the association between disease incidence (Z_1) and exposure to a suspected risk factor (Z_2) in the 2×2 table of individuals classified as cases or controls and as exposed or non-exposed (Breslow, 1976). One of the advantages of using the odds ratio as the measure of the strength of the association is that it is invariant under sampling designs that depend on the marginal characteristics of the variables (Chen, 2007). In testing and measuring the associations with discrete variables, the odds-ratio parameter is often adopted (Breslow, 1981, 1996; Liang, 1985; Hanfelt and Liang, 1998). Modeling the heterogeneity of the odds ratio for binary outcomes parametrically has received a lot of attention (see, Zelen, 1971; Breslow, 1976; Davis, 1985; Liang, 1985; Liang et al., 1986; Hanfelt and Liang, 1998). However, at times, a parametric regression model may create a very large modeling bias and be a misspecification, which may lead to biased and inefficient estimators (see, Fan and Gijbels, 1996), thus, we consider in this paper the nonparametric model for the odds ratio.

For the 2×2 table at point T_i (see, Table 1), suppose that the probabilities of the four cells are $p_{11}(T_i)$, $p_{12}(T_i)$, $p_{21}(T_i)$, and $p_{22}(T_i)$, respectively, and the log odds ratio is $g(T_i)$, for $i = 1, 2, \ldots, n$, where $p_{11}(\cdot)$, $p_{12}(\cdot)$, $p_{21}(\cdot)$, and $p_{22}(\cdot)$ are unknown positive regression functions, satisfying $p_{11}(t) + p_{12}(t) + p_{21}(t) + p_{22}(t) = 1$ for each t, and the log odds-ratio function $g(t)$, i.e., $\log\{(p_{11}(t)p_{22}(t))/(p_{12}(t)p_{21}(t))\}$, is also unknown. This paper is concerned with the smoothing problem of the log odds-ratio function for sparse data, where the number of tables at all points is large, and the table sizes are all small. There exist many researches concerning parametric regression analysis of odds ratio for sparse data. For example, by adopting the noncentral hypergeometric distribution, Breslow (1976) developed an iterative procedure to estimate the regression parameters based on a simply expressed likelihood equation; Hanfelt and Liang (1998) used the Mantel–Haenszel estimating function and the Mantel–Haenszel quasi-likelihood function to get the estimates. Nevertheless, to our knowledge, little literature about the nonparametric regression of odds ratio exists.

We organize our article as follows. In Section 2, we derive two estimators of the unknown log odds-ratio function. The asymptotic properties of the estimators are also presented. In Section 3, simulation studies are conducted to evaluate the

* Corresponding author.
E-mail address: chenzq453@nenu.edu.cn (Z. Chen).

0167-7152/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2011.06.017

Table 1
2×2 table at T_i.

	No disease	Disease
Controls	n_{11i}	n_{12i}
Cases	n_{21i}	n_{22i}

performances of the proposed methods. A brief discussion is presented in Section 4. Technical conditions and a lemma are relegated to the Appendix.

2. New estimation approaches

2.1. Table-wise equation based approach

In the log odds-ratio parametric regression problem, i.e., $m_i = \exp(T_i\beta)$, a Mantel–Haenszel estimating function for regression parameter β was constructed by Hanfelt and Liang (1998) as

$$h(\text{data};\beta) := \sum_{i=1}^{n} \frac{\partial m_i}{\partial \beta}\omega_i(m_i)(n_{11i}n_{22i} - m_i n_{12i}n_{21i}), \tag{1}$$

$\hat{\beta}$ satisfies $h(\text{data};\hat{\beta}) = 0$, and is a consistent estimator of β. The weights $\omega_i(m_i)$ are selected to be $m_i(\beta)^{-3/2}N_i^{-1}$ to minimize the asymptotic conditional variance of $\hat{\beta}$ under the naive assumption of no association between incidence and exposure, and to enforce the interchangeability property, where $N_i = n_{11i} + n_{12i} + n_{21i} + n_{22i}$, which is the count of events at T_i (see also, Hanfelt and Liang, 1998). In the simple case for $m_i = m$, for $i = 1, 2, \ldots, n$, i.e., the common odds-ratio case, $\hat{m} = \sum_{i=1}^{n}(n_{11i}n_{22i}/N_i)/\sum_{i=1}^{n}(n_{12i}n_{21i}/N_i)$, which is shown by Breslow (1981) to be computationally efficient, reasonable, and as efficient as the estimator based on the noncentral hypergeometric distribution (Cox, 1970) with asymptotic variance being the criterion.

When the log odds-ratio function is set to be $g(t)$, motivated by Mantel–Haenszel estimating function (1), based on kernel regression techniques, we now define the so-called table-wise equation,

$$L_1\{\text{data}; m(t)\} := \sum_{i=1}^{n} K_h(T_i - t)\{n_{11i}n_{22i} - m(t)n_{12i}n_{21i}\}/N_i,$$

to find the estimator of $g(t)$, where $K_h(t) = h^{-1}K(t/h)$ with $K(\cdot)$ being a kernel function, and $m(t) = \exp\{g(t)\}$. By solving $L_1\{\text{data}; m(t)\} = 0$, we can obtain the estimator $\hat{m}(t)$ of $m(t)$, which is

$$\sum_{i=1}^{n} K_h(T_i - t)(n_{11i}n_{22i}/N_i) \Big/ \sum_{i=1}^{n} K_h(T_i - t)(n_{12i}n_{21i}/N_i).$$

That estimator parallels the estimator given by Breslow (1981). We then get the table-wise equation based estimator of $g(t)$ (TWEE), denoted as $\hat{g}_{\text{TW}}(t)$, which is

$$\hat{g}_{\text{TW}}(t) = \log\left\{\sum_{i=1}^{n} K_h(T_i - t)(n_{11i}n_{22i}/N_i) \Big/ \sum_{i=1}^{n} K_h(T_i - t)(n_{12i}n_{21i}/N_i)\right\}.$$

Define

$$\hat{\psi}_1(t) := \sum_{i=1}^{n} K_h(T_i - t)(n_{11i}n_{22i}/N_i) \Big/ \sum_{i=1}^{n} K_h(T_i - t)(N_i - 1),$$

and

$$\hat{\psi}_2(t) := \sum_{i=1}^{n} K_h(T_i - t)(n_{12i}n_{21i}/N_i) \Big/ \sum_{i=1}^{n} K_h(T_i - t)(N_i - 1).$$

Suppose that $N_i \neq 1$, for each $i \in \{1, 2, \ldots, n\}$, then, it can be derived easily that $\hat{\psi}_1(t)$ and $\hat{\psi}_2(t)$ are consistent estimators of $p_{11}(t)p_{22}(t)$ and $p_{12}(t)p_{21}(t)$, respectively. Hence, $\hat{g}_{\text{TW}}(t)$ is a consistent estimator of $g(t)$.

Theorem 2.1. *Suppose that the assumptions in the Appendix hold, and for each $i \in \{1, 2, \ldots, n\}$, $N_i \neq 1$. If $nh^5 = O(1)$ as $n \to \infty$, then, for t an interior point of Ω,*

$$\sqrt{Nh}(\hat{g}_{\text{TW}}(t) - g(t) - h^2 c_2(K)v(t)) \xrightarrow{L} N\left(0, \frac{\gamma_0(K)\sigma(t)}{f(t)}\right),$$

with $v(t)$ and $\sigma(t)$ defined in the Appendix, where $N = \sum_{i=1}^{n} N_i$, $c_2(K) = \int t^2K(t)\mathrm{d}t$, and $\gamma_0(K) = \int K^2(t)\mathrm{d}t$.

This theorem is a direct result from Lemma 1 in the Appendix using the delta method.

2.2. Cross-table-wise equation based method

If $m(t)$ is the true parameter, we can show an important feature that is

$$E[r_{ij}\{m(t)\}] = O(h) = o(1),$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$, where $r_{ij}\{m(t)\} = K_h(T_i - t)K_h(T_j - t)\{n_{11i}n_{22j} - m(t)n_{12i}n_{21j}\}$. By assigning probability mass $1/n^2$ to each cross-table function $r_{ij}\{m(t)\}$, an empirical estimator of $E[r_{ij}\{m(t)\}]$, i.e., $\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\{m(t)\}/n^2$, is obtained, then, we define the so-called cross-table-wise equation as

$$L_2(\text{data}; m(t)) := \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{r_{ij}\{m(t)\}}{n^2}.$$

Solve $L_2(\text{data}; m(t)) = 0$, the estimator $\hat{m}(t)$ of $m(t)$ can be obtained, which is

$$\left\{\sum_{i=1}^{n} K_h(T_i - t)n_{11i}\sum_{i=1}^{n} K_h(T_i - t)n_{22i}\right\} \Big/ \left\{\sum_{i=1}^{n} K_h(T_i - t)n_{12i}\sum_{i=1}^{n} K_h(T_i - t)n_{21i}\right\}.$$

Then, the cross-table-wise equation based estimator of $g(t)$ (CTWEE), denoted as $\hat{g}_{\text{CTW}}(t)$, can be obtained, which is

$$\log\left[\left\{\sum_{i=1}^{n} K_h(T_i - t)n_{11i}\sum_{i=1}^{n} K_h(T_i - t)n_{22i}\right\} \Big/ \left\{\sum_{i=1}^{n} K_h(T_i - t)n_{12i}\sum_{i=1}^{n} K_h(T_i - t)n_{21i}\right\}\right].$$

We now state the intuitive interpretation for $\hat{g}_{\text{CTW}}(t)$. Motivated by the idea of nonparametric regression techniques, we give the local average 2×2 table at t in the interior of Ω as

$$\sum_{i=1}^{n} K_h(T_i - t)\text{Table}(T_i) \Big/ \sum_{i=1}^{n} N_i K_h(T_i - t), \tag{2}$$

where $\text{Table}(T_i)$ denotes the 2×2 table at T_i. We claim that the (i, j)th cell of (2), denoted as $\hat{p}_{ij}(t)$, is a consistent estimate of $p_{ij}(t)$, for $i = 1, 2;\ j = 1, 2$. $\hat{g}_{\text{CTW}}(t)$ is actually $\log[\{\hat{p}_{11}(t)\hat{p}_{22}(t)\}/\{\hat{p}_{12}(t)\hat{p}_{21}(t)\}]$, consequently, the consistence of $\hat{g}_{\text{CTW}}(t)$ can be got. Intuitively, taking the kernel to be the uniform kernel, $\hat{g}_{\text{CTW}}(t)$ is actually the sample log odds ratio of the table

$$\sum_{|T_i - t| \le 0.5h} \text{Table}(T_i),$$

which is approximate to that by sampling repeatedly infinite times, i.e., $O(nh)$ times, at t, thus the consistency of $\hat{g}_{\text{CTW}}(t)$ is guaranteed, as $n \to \infty$, under Assumptions 2 and 3 in the Appendix.

Remark 1. The positiveness of $\hat{p}_{11}(t)$, $\hat{p}_{12}(t)$, $\hat{p}_{21}(t)$, and $\hat{p}_{22}(t)$, and $\hat{p}_{11}(t) + \hat{p}_{12}(t) + \hat{p}_{21}(t) + \hat{p}_{22}(t) = 1$ are guaranteed.

Let $\hat{P}(t) := (\hat{p}_{11}(t), \hat{p}_{12}(t), \hat{p}_{21}(t), \hat{p}_{22}(t))'$, $P(t) := (p_{11}(t), p_{12}(t), p_{21}(t), p_{22}(t))'$, $B(t) := (b_{11}(t), b_{12}(t), b_{21}(t), b_{22}(t))'$ with $b_{ij}(t) = \frac{1}{2}p_{ij}^{(2)}(t) + \frac{f^{(1)}(t)p_{ij}^{(1)}(t)}{f(t)}$, for $i = 1, 2;\ j = 1, 2$, $\text{diag}\{P(t)\}$ is a diagonal matrix with the elements of $P(t)$ on the main diagonal, and $l(t) := b_{11}(t)/p_{11}(t) - b_{12}(t)/p_{12}(t) - b_{21}(t)/p_{21}(t) + b_{22}(t)/p_{22}(t)$.

Theorem 2.2. *Suppose that the assumption s given in the Appendix hold. We have the following results:*

If $nh^5 = O(1)$ as $n \to \infty$, then,

$$\sqrt{Nh}(\hat{P}(t) - P(t) - h^2 c_2(K)B(t)) \xrightarrow{L} N\left(0, \frac{\gamma_0(K)}{f(t)}[\text{diag}\{P(t)\} - P(t)P(t)']\right),$$

and

$$\sqrt{Nh}\left(\hat{g}_{\text{CTW}}(t) - g(t) - h^2 c_2(K)l(t)\right) \xrightarrow{L} N\left(0, \frac{\gamma_0(K)\delta(t)}{f(t)}\right),$$

for every t in the interior of Ω, where $N = \sum_{i=1}^{n} N_i$, and $\delta(t) = \frac{1}{p_{11}(t)} + \frac{1}{p_{12}(t)} + \frac{1}{p_{21}(t)} + \frac{1}{p_{22}(t)}$.

Following the same lines as the proof of Lemma 1, we can get the first equation in the theorem above, and applying the delta method, the second one can be verified.

Remark 2. The number of tables falling into the neighborhood of t with bandwidth h is $O(Nh)$, thus, as is stated in the theorem, the convergence rate of the CTWEE is $\sqrt{Nh}$, intuitively.

Table 2
Summary results of the Monte Carlo bias and mean squared error at 200 fixed points based on 500 simulated datasets. Entries from up to down are the sample quartiles from the 200 biases and mean squared errors for $n = 100$, 150 and 300. MSE: mean squared error; Q1: lower quartile; Q3: upper quartile.

		TWEE		CTWEE		Parametric	
		Bias	MSE	Bias	MSE	Bias	MSE
$n = 100$	Q1	−0.1400	0.0331	−0.1393	0.0322	−0.1791	0.0434
	Median	−0.0029	0.0412	−0.0038	0.0400	−0.0144	0.0602
	Q3	0.1062	0.0515	0.1018	0.0503	0.1700	0.0748
$n = 150$	Q1	−0.1146	0.0237	−0.1128	0.0230	−0.1762	0.0316
	Median	−0.0018	0.0300	−0.0012	0.0291	−0.0149	0.0509
	Q3	0.0935	0.0353	0.0898	0.0345	0.1748	0.0680
$n = 300$	Q1	−0.0870	0.0130	−0.0850	0.0127	−0.1826	0.0204
	Median	−0.0004	0.0167	−0.0004	0.0162	−0.0155	0.0412
	Q3	0.0664	0.0196	0.0653	0.0192	0.1710	0.0609

3. Simulation study

In this section, we investigate the finite sample performances of the proposed estimators in Section 2. We generate 500 datasets, each consisting of n points at which the 2×2 tables are observed. We take n = 100, 150 and 300 to represent small, moderate and large numbers of tables, respectively. The T_i, $i = 1, 2, \ldots, n$, are generated from a uniform distribution on the interval [0, 1]. Given t, we take the log odds-ratio function $g(t) = \cos(2\pi t)$, the expectation of the binary variables $Z_1(t)$ and $Z_2(t)$ are set to be $e^t/(1 + e^t)$ and 0.6, respectively. The number of observations N_i for table i is set to be 25, for $i = 1, 2, \ldots, n$. From the log odds-ratio and marginal mean functions, i.e., $g(t)$, $EZ_1(t)$, and $EZ_2(t)$, the iterative proportional fitting procedure (see also, Fitzmaurice and Laird, 1993) can be applied to obtain the probabilities of the four cells, i.e., $p_{11}(t)$, $p_{12}(t)$, $p_{21}(t)$, and $p_{22}(t)$. We generate the cell counts of the table at T_i, $(n_{11i}, n_{12i}, n_{21i}, n_{22i})$, from a multinomial $(N_i; p_{11}(T_i), p_{12}(T_i), p_{21}(T_i), p_{22}(T_i))$ distribution, for $i = 1, 2, \ldots, n$. Here, our simulated datasets are sparse datasets. Our main focus is the estimation of the log odds-ratio function. For comparison purpose, we give the parametric model for the log odds ratio, i.e., $g(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3$, and use the Mantel–Haenszel estimating function (1) to estimate $\beta := (\beta_1, \beta_2, \beta_3, \beta_4)'$. If we get $\hat{\beta}$, $\hat{g}(t)$ under the parametric model can be obtained. The three estimators are evaluated, the TWEE, the CTWEE and the estimator under the parametric model. Since the T vary from dataset to dataset, estimates are obtained at 200 fixed equally spaced grid points within the range of T throughout the simulation. Numerical properties of the three estimates are then investigated.

In the simulation study, the Monte Carlo bias and mean squared error (MSE) for each estimator are obtained at each of the 200 fixed points. Table 2 summarizes the results. The MSE of the TWEE is similar to that of the CTWEE, which indicates that the table-wise equation based method is as efficient as the cross-table-wise equation based method when MSE is used as the criterion. The two proposed estimators show little biases. Compared to the estimator under the parametric model, both proposed estimators exhibit less biases and smaller MSEs regardless of the number of tables for each dataset (i.e., n). This is due to the fact that the parametric model is a misspecification.

We now compare the performances of two estimators through averages over all simulated datasets. Here, an index R is calculated for comparing both proposed estimators with the estimator under the parametric model for each dataset. We obtain the sums of squared deviations between the estimated and the true g values at the 200 fixed points for all estimators. We define R to be the ratio of the two sums of squared deviations, with the numerator again calculated using the estimator under the parametric model, and the denominator using the new estimator. The sample quantile plot of R is given in Fig. 1 for $n = 150$. An R larger than 1 indicates that the proposed estimator has a smaller sum of squared errors than the estimator under the parametric model. It is clearly showed by Fig. 1 that the proposed estimators are both superior to the estimator under the parametric model for almost all 500 datasets. This is also due to the fact that the parametric model is a misspecification. Furthermore, by the closeness of the solid and short dashes curves, we observe that the TWEE behaves approximately as well as the CTWEE.

4. Discussion

Motivated by the Mantel–Haenszel estimating function (Hanfelt and Liang, 1998), we propose the table-wise equation based estimator for the log odds-ratio function for sparse data. We also propose the cross-table-wise equation based estimator. Both estimators are consistent and attain asymptotic normality. The two estimators are computational straightforwardly from the samples. As is indicated in the simulation study, both proposals outperform the estimator based on the Mantel–Haenszel estimating function (1) under the parametric model, and the two proposals enjoy similar efficiency.

Acknowledgments

We are grateful to two anonymous referees, an associate editor and the co-editor for helpful and constructive comments on earlier versions, which led to significant improvements. This work has been partly supported by Program for New Century

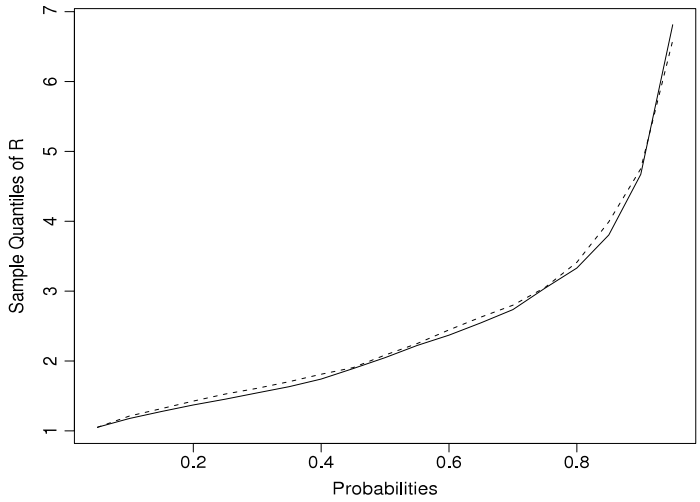

Fig. 1. Sample quantile plot of R for the table-wise equation based (solid) and the cross-table-wise equation based (short dashes) estimators versus the estimator under the parametric model among the 500 simulated datasets with $n = 150$.

Excellent Talents in University, National Nature Science Foundation of China (NO. 11071035), National Nature Science Foundation of China (NO. 10931002) and the Fundamental Research Funds for the Central Universities (NO. 09SSXT116).

Appendix

We investigate the large sample properties of the estimators given in Section 2. For this purpose, we give the following regular conditions. They may not be the weakest possible conditions, but they are imposed to facilitate the proofs.

Assumptions:

1. $T_1, \ldots, T_n$ are independently and identically sampled from a density having a version $f(\cdot)$ with compact support Ω. In addition, f is twice continuously differentiable, and is bounded away from 0 in a neighborhood of the each t belonging to the interior of Ω. The function $K(\cdot)$ is a symmetric density function.

2. The log odds-ratio function $g(t)$, and the marginal mean functions, i.e., $EZ_1(t)$ and $EZ_2(t)$, have continuous second order derivatives in a neighborhood of each t belonging to the interior of Ω.

3. $h \to 0$ and $nh^3 \to \infty$, as $n \to \infty$.

4. The tables at all points are independent, and the cell counts of the table at $T_i(n_{11i}, n_{12i}, n_{21i}, n_{22i})$ have a multinomial $(N_i; p_{11}(T_i), p_{12}(T_i), p_{21}(T_i), p_{22}(T_i))$ distribution, for $i = 1, 2, \ldots, n$.

5. There exists an integer K_0 such that $N_i \le K_0$, for $i = 1, 2, \ldots, n$.

Define $\Sigma(t) := \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_{22}(t) \end{pmatrix}$, with $\sigma_{11}(t)$ being $\lim_{n\to\infty}[N\{(N-3n+2M)p_{11}(t)\omega_1(t)+(N-3n+2M)p_{22}(t)\omega_1(t)+(n-M)\omega_1(t)-2(2N-5n+3M)\omega_1^2(t)\}/(N-n)^2]$, $\sigma_{22}(t)$ being $\lim_{n\to\infty}[N\{(N-3n+2M)p_{12}(t)\omega_2(t)+(N-3n+2M)p_{21}(t)\omega_2(t)+(n-M)\omega_2(t)-2(2N-5n+3M)\omega_2^2(t)\}/(N-n)^2]$, and $\sigma_{12}(t)$ being $\lim_{n\to\infty}[-2N(2N-5n+3M)\omega_1(t)\omega_2(t)/(N-n)^2]$, where $M = \sum_{i=1}^{n}(1/N_i)$, $\omega_1(t) = p_{11}(t)p_{22}(t)$, and $\omega_2(t) = p_{12}(t)p_{21}(t)$. Let $\omega(t) := (\omega_1(t), \omega_2(t))'$; $c(t) := (c_1(t), c_2(t))'$ with $c_i(t)$ being $\omega_i^{(1)}(t)f^{(1)}(t)/f(t) + \frac{1}{2}\omega_i^{(2)}(t)$, for $i = 1, 2$; $v(t) := c_1(t)/\omega_1(t) - c_2(t)/\omega_2(t)$; and $\sigma(t) := \sigma_{11}(t)/\omega_1^2(t) - 2\sigma_{12}(t)/\{\omega_1(t)\omega_2(t)\} + \sigma_{22}(t)/\omega_2^2(t)$.

Lemma 1. *Suppose that the assumptions given above hold, and $N_i \ne 1$, for each $i \in \{1, 2, \ldots, n\}$. If $nh^5 = O(1)$ as $n \to \infty$, then, for t an interior point of Ω,*

$$\sqrt{Nh}\left(\hat{\psi}(t) - \omega(t) - h^2 c_2(K)c(t)\right) \xrightarrow{L} N\left(0, \frac{\gamma_0(K)}{f(t)}\Sigma(t)\right),$$

where $\hat{\psi}(t) = (\hat{\psi}_1(t), \hat{\psi}_2(t))'$.

Proof. $E(n_{11i}n_{22i} \mid T_i) = N_i(N_i-1)p_{11}(T_i)p_{22}(T_i)$, and $\mathrm{Var}(n_{11i}n_{22i} \mid T_i) = \{-2N_i(N_i-1)(2N_i-3)p_{11}^2(T_i)p_{22}^2(T_i) + N_i(N_i-1)(N_i-2)p_{11}^2(T_i)p_{22}(T_i)+N_i(N_i-1)(N_i-2)p_{11}(T_i)p_{22}^2(T_i)+N_i(N_i-1)p_{11}(T_i)p_{22}(T_i)\}$. Also, $E(n_{12i}n_{21i} \mid T_i)$ and $\mathrm{Var}(n_{12i}n_{21i} \mid T_i)$ can be expressed in the similar forms. Apply the familiar proof techniques in Fan and Gijbels (1996, Chapter 5) or Pagan and Ullah (1999, Chapter 3), the result follows. □

References

Breslow, N., 1976. Regression analysis of the log odds ratio: a method for retrospective studies. Biometrics 32, 409–416.
Breslow, N., 1981. Odds ratio estimators when the data are sparse. Biometrika 68, 73–84.

Breslow, N., 1996. Statistics in epidemiology: the case-control study. J. Amer. Statist. Assoc. 91, 14–28.
Chen, H.Y., 2007. A semiparametric odds ratio model for measuring association. Biometrics 63, 413–421.
Cox, D.R., 1970. The Analysis of Binary Data. Methuen, London.
Davis, L.J., 1985. Generalization of the Mantel–Haenszel estimator to nonconstant odds ratios. Biometrics 41, 487–495.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.
Fitzmaurice, G.M., Laird, N.M., 1993. A likelihood-based method for analysing longitudinal binary responses. Biometrika 80, 141–151.
Hanfelt, J.J., Liang, K.Y., 1998. Inference for odds ratio regression models with sparse dependent data. Biometrics 54, 136–147.
Liang, K.Y., 1985. Odds ratio inference with dependent data. Biometrika 72, 678–682.
Liang, K.Y., Beaty, T.H., Cohen, B.H., 1986. Application of odds ratio regression models for assessing familial aggregation from case-control studies. Am. J. Epidemiol. 124, 678–683.
Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press.
Zelen, M., 1971. The analysis of several 2×2 contingency tables. Biometrika 58, 129–137.