
ANALYSIS OF A SEQUENCE OF DEPENDENT 2 × 2 TABLES

B. C. KOU
Department of IEOR, 510 Wold Building, Columbia University
530 West 120th Street, New York, NY 10027, U.S.A.

ZHIHONG YING
Department of Statistics, MC 4250, Columbia University
1255 Amsterdam Avenue, 10th Floor, New York, NY 10027, U.S.A.

A sequence of dependent 2 × 2 contingency tables often arises in epidemiologic cohort studies, controlled clinical trials, and other follow-up studies. Due to dependence, it is unclear, however, whether and how the conditional approach for a single 2 × 2 table can be extended to analyze a sequence of dependent 2 × 2 tables. We show that distributional properties can be derived by considering a "tangent" sequence of independent 2 × 2 tables, with each 2 × 2 table being represented by a sum of independent, yet not identically distributed, Bernoulli trials, whose success probabilities can easily be computed (Kou and Ying, 1998). The method has four applications: (1) We provide a characterization of the validity of a weighted log-rank test. (2) The method leads to a simple algorithm to compute the maximum partial likelihood estimator of the common odds ratio, as well as its variance estimator. The efficiency over the traditional Mantel-Haenszel estimator is also demonstrated. (3) We show how to use the method to provide estimation in Bradlow-Prentice model for regression with nonhomogeneous odd ratios. (4) The method is applied to analysis of a proportional hazards model with tied observations. The computation is straightforward by using a link with the roots of Jacobi polynomials.

1. Introduction

Suppose two coins with success probabilities \(p_1\) and \(p_2\) are tossed \(N_1\) and \(N_2\) times. Let \(M_1\) and \(M_2\) be the numbers of heads and tails, respectively, in the \(N = N_1 + N_2\) tosses, and \(X\) the number of heads from the first coin. It is known that the conditional distribution of \(X\) given the \(M_1\) and \(M_2\) is noncentral hypergeometric

\[
P(X = x) = \frac{\binom{N_1}{M_1} \binom{N_2}{M_2} p^x (1-p)^{N-x}}{\sum_{x=0}^{N} \binom{N_1}{M_1} \binom{N_2}{M_2} p^x}, \quad 0 \leq x \leq N, \tag{1.1}
\]

where \(p = \frac{p_1}{p_1 + p_2}\).
where \( \theta = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \) is the odds ratio parameter, \( L = \max(0, M_1 - M_2) \) and \( S = \min(N_1, M_1) \); see Breslow and Day (1980, p.125).

In particular, if \( p_1 = p_2 \), then \( \theta = 1 \) and (1.1) reduces to the (central) hypergeometric distribution

\[
P(X = x) = \binom{N_1}{x} \binom{N_2}{M_1 - x} / \binom{N}{M_1}, \quad L \leq x \leq S,
\]

(1.2)
giving rise to the celebrated Fisher’s exact test for the hypothesis \( p_1 = p_2 \).

The noncentral hypergeometric family (1.1) is frequently used in epidemiological studies to investigate relationships between occurrence of a disease and exposure to a possible risk factor. Such studies may be summarized into a 2 x 2 table (see below), in which \( N_1 \) and \( N_2 \) are taken as the numbers of persons in the exposed and the unexposed groups, respectively, and \( M_1 \) and \( M_2 \) the numbers of diseased and disease-free individuals, respectively (Breslow and Day, 1980, p.124).

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Disease-Free</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>( M_1 - X )</td>
<td>( N_1 )</td>
</tr>
<tr>
<td>Unexposed</td>
<td>( X )</td>
<td>( M_2 )</td>
</tr>
</tbody>
</table>

The hypothesis that the disease rates for the two groups, the exposed and the unexposed, are the same is therefore tantamount to \( \theta = 1 \), and Fisher’s exact test applies. The same table arises in controlled clinical experiments as well, where the exposure factor becomes the treatment/control indicator.

Of pivotal concern in the current paper is a sequence of \( K \) dependent 2 x 2 tables, where \( K \) may be a random variable or stopping time, and the dependence structure among the tables may not be fully known (e.g. due to possible censorship in the data). One motivation comes from epidemiological cohort studies, in which individuals are identified along with their exposure history, and are followed forward in time to ascertain the occurrences of the diseases of interest so that the exposure information can be related to subsequent disease experience (Breslow and Day, 1987). Such studies are useful to establish causality between exposure to a possible risk factor and occurrence of a disease. Stratification in time is often needed, giving rise to a sequence of dependent 2 x 2 tables. Such dependent sequence also appears in controlled clinical trials, where patients receiving treatment/control are followed in time until occurrence of certain clinical endpoints. See Miller (1981, Chapter 4) for an excellent introduction on how such sequences of dependent 2 x 2 tables arise from medical follow-up studies.

Due to the nature of follow-up studies, the 2 x 2 tables thus constructed are typically dependent in such a way that the conditional distribution of the 4th table given its margin in noncentral hypergeometric form (1.1).

This property allows us to make use of the decoupling method (Kwapień and Woyczyński, 1992; de la Peña and Giné, 1999) to connect the sequence of such dependent 2 x 2 tables to a “tangent” sequence of independent tables. In conjunction with a representation that a noncentral hypergeometric random variable may be expressed as a sum of independent Bernoulli random variables (Kou and Ying, 1996), the decoupling method enables us to establish validity of normal approximations to weighted sums of possibly dependent 2 x 2 tables.

Our method has implications in several aspects. First, it allows us to rigorously investigate the weighted log-rank statistic to test the hypothesis that the odds ratio is equal to one, i.e. no treatment effect. In the existing literature, the asymptotic theory for the log-rank statistic applies only to the situation of a large collection of small tables or a small number of large tables. We establish here a new asymptotic property under a minimal condition that the total conditional variance goes to infinity, no matter how many large or small tables are. See Section 3.

Second, our approach provides a simple and efficient way to compute the maximum (partial) likelihood estimator of the common odds ratio. A widely used estimator for the common odds ratio for a sequence of independent 2 x 2 tables is due to Mantel and Haenszel (1958). However, complication could arise in estimating its variance when the tables are a mixture of large and small frequencies (Robins, Breslow and Greenland, 1986; Phillips and Holland, 1987). For a review of the Mantel-Haenszel method, see Breslow (1996). Alternatively, one may use the likelihood approach to estimate the odds ratio. Assuming the tables to be independent, the conditional likelihood given all the margins is a product of hypergeometric probability mass functions of form (1.1). When all tables are independent, it is possible, though computationally demanding, to implement certain exact inference procedures (Cytel Inc., 2003), or use saddle point approximation (Strawderman and Wells, 1998). However, this approach obviously fails when the independent assumption is not valid. Our method leads to a fast way to compute the maximum partial likelihood estimator as well its asymptotic variance via a connection with the roots of Jacobi polynomials. This simplifies the computation effort significantly. The calculation is especially useful in survival analysis where the dependence among tables arises naturally. See Section 4.
words, they are the same for both hypergeometric and noncentral hypergeometric distributions. This fact greatly reduces computational burden in dealing with the noncentral hypergeometric distribution.

Specifically, for the 4th table, the roots $\lambda_k^{(\theta)}$ are simple functions of the roots of the Jacobian polynomials, which are widely used in mathematics and engineering literature (Kos and Ying, 1990). The roots can be determined easily by using software packages such as Mathematica (Wolfram, 1991). Alternatively, one can utilize the associated matrix of which $\theta$ is the characteristic polynomial. Writing the probability generating function (2.3) as $\Phi(x) = \sum_{n=0}^{\infty} a_n x^n \left( a_0 \neq 0 \right)$, its roots are exactly the eigenvalues of its associate matrix

$$
Q = \begin{pmatrix}
\frac{\theta - 1}{\theta - 1} & 0 & \dots & 0 \\
0 & \frac{\theta - 1}{\theta - 1} & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \frac{\theta - 1}{\theta - 1}
\end{pmatrix}
$$

See Johnson and Riess (1982, Section 4.4.3). Again, the eigenvalues can be determined easily by the existing software packages. For example, a relevant command in Splus (MathSoft, 1995 is eigval(Q)) values for obtaining eigenvalues of $Q$.

From (2.2), it follows that the mean and variance of $X$ can be expressed in terms of $\lambda_k$ and $\theta$:

$$
E(x) = \sum_{i=1}^{\theta} \frac{1}{1 + \theta - 1} = \frac{\theta^{(1)}}{\theta - 1} \text{Var}(X) = \sum_{i=1}^{\theta} \theta^{(1)} = \frac{\theta^{(1)}}{\theta - 1},
$$

where the subscript $\theta$ in $E$ and $\text{Var}$ indicates that the expectation and variance are taken with $\theta$ being the true parameter.

2.3. Decoupling method

The second key element in our analysis is the "decoupling" method, a recent development in probability theory. A systematic exposure of the method can be found in Kwapień and Woyczyński (1992) and de la Peña and Giné (1999). To apply the method, we need to introduce two definitions, which are followed by a lemma.

Definition 2.1. Two sequences of random variables, $(\theta_j, k \geq 1)$ and $(\delta_k, k \geq 1)$, are said to be tangent with respect to $(\theta_k)$ if, for each $k$,
and that 

\[ \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = 0 \text{ in probability.} \]

Then \( U \) defined by (2.1) converges to distribution \( P_{0}(0) \) for every \( \alpha \). Exchanging \( X \) with \( X_{(1)} \), (2.2) is satisfied by (2.1) and (2.3).

Theorem 2.3. Suppose that there exist a constant \( 0 < \alpha \leq 1 \) such that

\[ \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = 0 \text{ in probability.} \]

Then (2.1) converges to distribution \( P_{0}(0) \) for every \( \alpha \). Exchanging \( X \) with \( X_{(1)} \), (2.2) is satisfied by (2.1) and (2.3).

Proof. Let \( X \) be a random variable with density \( f_X(x) \) and let \( U \) be a random variable with density \( f_U(u) \). By Lemma 2.1, (2.1) satisfies the condition for \( U \) to converge to distribution \( P_{0}(0) \) as \( n \to \infty \).

Theorem 2.3. Suppose that there exist a constant \( 0 < \alpha \leq 1 \) such that

\[ \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = 0 \text{ in probability.} \]

Then (2.1) converges to distribution \( P_{0}(0) \) for every \( \alpha \). Exchanging \( X \) with \( X_{(1)} \), (2.2) is satisfied by (2.1) and (2.3).

Proof. Let \( X \) be a random variable with density \( f_X(x) \) and let \( U \) be a random variable with density \( f_U(u) \). By Lemma 2.1, (2.1) satisfies the condition for \( U \) to converge to distribution \( P_{0}(0) \) as \( n \to \infty \).
nal. Therefore it appears that \( U \) should always be approximately normal, no matter how the tables are arranged. Yet neither the representation by independent Bernoulli random variables nor the martingale central limit theorem can be applied directly. In these respects, Theorem 3.1 presented here becomes an effective tool, complementing the available methods.

4. Maximum Partial Likelihood Estimator for the Common Odds Ratio

In many applications, especially medical follow-up studies, it is common to assume homogeneity of the odds ratio parameters, i.e., \( \theta = \theta^* \) (Mantel and Haenszel, 1959; Breslow, 1960). Therefore, an important statistical problem is to estimate the common odds ratio \( \theta \) and that will be the concern of this section.

4.1. The estimator

The conditional probability mass function of the \( k \)-th table given its margins and the \( k - 1 \) preceding tables can be written as

\[
\pi_k(M_k^N, M_k^0, N_k^0, N_k^0) = \frac{\binom{N_k^0}{x_k^0} \binom{N_k^0}{y_k^0} \binom{N_k^0}{z_k^0}}{\sum_{a_k^0, b_k^0, c_k^0} \binom{N_k^0}{a_k^0} \binom{N_k^0}{b_k^0} \binom{N_k^0}{c_k^0}},
\]

where \( \pi_k \) is the conditional probability mass of \( (M_k^N, M_k^0, N_k^0, N_k^0) \) given the \( k - 1 \) preceding tables. Following Cox (1975), we ignore the \( \pi_k \)'s and obtain the following partial likelihood

\[
\prod_{k=1}^{K} \frac{\binom{N_k^0}{x_k^0} \binom{N_k^0}{y_k^0} \binom{N_k^0}{z_k^0}}{\sum_{a_k^0, b_k^0, c_k^0} \binom{N_k^0}{a_k^0} \binom{N_k^0}{b_k^0} \binom{N_k^0}{c_k^0}}.
\]

Notice that (4.1) becomes the full likelihood when the tables are independent and their margins are fixed. By taking the logarithm of the partial likelihood and then setting its derivative with respect to \( \theta \) equal to zero, we see that the maximum partial likelihood estimator (MPLE), denoted by \( \hat{\theta} \), satisfies the following equation

\[
\sum_{k=1}^{K} \left[ X_k^0 - E_k^{(k-1)} X(k) \right] = 0,
\]

which, through (2.4), is the same as

\[
\sum_{k=1}^{K} X_k^0 - \frac{X(k)}{\theta - 1} \frac{1}{\theta - 1} \frac{1}{\theta - 1} = 0.
\]

The simple form of (4.2) allows the characterization of existence and uniqueness of the MPLE.

Proposition 4.1. A necessary and sufficient condition that guarantees the existence and uniqueness of the MPLE \( \hat{\theta} \) is

\[
\sum_{k=1}^{K} L_k^0 \leq \sum_{k=1}^{K} X_k^0 < \sum_{k=1}^{K} S_k^0,
\]

since \( L_k^0 \leq X_k^0 \leq S_k^0 \) for all \( k \), it follows that (4.3) is equivalent to that there exist \( k \) and \( k' \) such that \( L_k^0 < X_k^0 < S_k^0 \).

Proof: We first show that (4.3) is sufficient. It is obvious that, on \((0, \infty)\), \((1 + \theta^{-1})^{-1} - 1\) is strictly increasing in \( \theta \), if \( \lambda_0^0 > 0 \). Therefore, \( \sum_{k=1}^{K} L_k^0 \left( 1 + \theta^{-1} \right)^{-1} \) goes to \( \sum_{k=1}^{K} L_k^0 \) as \( \theta \to 0 \), and goes to \( \sum_{k=1}^{K} S_k^0 \) as \( \theta \to \infty \), because for each \( k \) there are exactly \( L_k^0 \neq 0 \) zero among \( X_k^0 \), \( 1 \leq i \leq S_k^0 \). We have then that, for any \( X = \sum_{k=1}^{K} L_k^0 \) or \( X = \sum_{k=1}^{K} S_k^0 \), whence the necessity is obtained as well.

It is not difficult to see that the partial likelihood estimation function is monotone decreasing and convex. Thus solving the MPLE \( \hat{\theta} \) in (4.2) is rather straightforward numerically (for example, by the standard Newton-Raphson method), once all \( \lambda_i^0 \) are calculated. The large sample properties of \( \hat{\theta} \) are summarized by the following theorem. Its proof is given in Appendix.

Theorem 4.1. Suppose that (5.8) is satisfied with \( \omega = 1 \). Then \( \hat{\theta} \) is consistent and asymptotically normal. More precisely, we have \( \hat{\theta} \to \theta \) in probability and

\[
\theta^{-1} \left( \sum_{k=1}^{K} \text{Var} \left( X(k) \right) \right)^{1/2} \left( \hat{\theta} - \theta \right) \to_L N(0,1),
\]

\[
\theta^{-1} \left( \sum_{k=1}^{K} \text{Var} \left( X(k) \right) \right)^{1/2} \left( \log \hat{\theta} - \log \theta \right) \to_L N(0,1).
\]
It is known that the estimator of the variance of a large number of large tables is consistent in the case of either a small number of large tables or a large number of small tables and is robust against non-normality of the errors. However, as in the case of the earlier estimators, the estimator is also computationally and conceptually simple.

A similar estimator is the sample variance estimator, which can be written as

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

where \(x_i\) is the sample mean and \(n\) is the sample size. The estimator is robust against non-normality and is computationally simple.

4.3. Comparison of estimators

Various other methods have been proposed for the inference of the common odds ratio parameter. Woolf (1955) suggested an estimand of the form

\[
\hat{\rho} = \frac{\chi^2}{2(\chi^2 + 1)}
\]

where \(\chi^2\) is the chi-squared statistic.

In conclusion, the estimators proposed in this section are robust against non-normality and are computationally simple. The choice of estimator depends on the specific application and the data at hand.
5. Analysis of the Breslow-Zelen Model

The homogeneity assumption about odds ratio parameters $\theta_k$ of the $K$ tables may be violated. To accommodate the possible nonhomogeneity, a useful regression model was proposed and studied by Zelen (1971) and Breslow (1970). Their model assumes relationship

$$\log(\theta_k) = \alpha + \beta \gamma_k, \quad 1 \leq k \leq K,$$

where $\gamma_k$ is a vector of covariates associated with the $k$th table, and $\alpha$ and $\beta$ the intercept and the regression parameters. Typically, the $\gamma_k$ include the stratification variable used to obtain the sequence of tables. See Breslow and Crowley (1984). Note that without including the $\gamma_k$, the model reduces to the setup of the preceding section with $\theta = e^\alpha$.

Our method yields a direct way for analyzing and computing the estimators for the Breslow-Zelen model. More precisely, the partial likelihood function may be obtained analogously to (4.1), with its $\theta$ replaced by $\exp(\alpha + \beta \gamma_k)$. Differentiating with respect to $\alpha$ and $\beta$, we arrive at the following estimating equation, which is analogous to (4.3),

$$\sum_{k=1}^{K} \frac{X_k(1) \cdot \sum_{i=1}^{n} \frac{1}{1 + \lambda_i} \cdot \exp(\alpha + \beta \gamma_{ki})}{\sum_{i=1}^{n} \frac{1}{1 + \lambda_i} \cdot \exp(\alpha + \beta \gamma_{ki})} (X_k - \sum_{i=1}^{n} \frac{1}{1 + \lambda_i} \cdot \exp(\alpha + \beta \gamma_{ki})) (X_k - \sum_{i=1}^{n} \frac{1}{1 + \lambda_i} \cdot \exp(\alpha + \beta \gamma_{ki})) = 0.$$  

(5.1)

The $\lambda_i$'s are the same as those in (4.2) and thus depend neither on the parameter values nor on the covariates.

Because the derivative matrix of left-hand side of (5.1) is negative definite, the log-likelihood function is concave so that numerically it is straightforward to compute MPL (x, $\hat{\beta}$) of ($\alpha$, $\beta$). The asymptotic variance-covariance matrix can be estimated quite easily by $\nabla^2 \chi^2(\cdot, \hat{\beta})$, where

$$\nabla I(\alpha, \beta) = \sum_{k=1}^{K} \left( \frac{\partial}{\partial \alpha} \left( \frac{1}{1 + \lambda_k} \cdot \exp(\alpha + \beta \gamma_k) \right) \right) \left( \frac{\partial}{\partial \beta} \left( \frac{1}{1 + \lambda_k} \cdot \exp(\alpha + \beta \gamma_k) \right) \right).$$

Suppose there exists a nonrandom sequence $\nu_k$ such that $\nabla^2 \chi^2(\cdot, \hat{\beta})$ converges to a nonrandom positive definite matrix. Then it is not difficult to show that ($\hat{\alpha}$, $\hat{\beta}$) is consistent and asymptotically normal under suitable normalization. Hence, inference procedures such as testing and interval estimation based on ($\hat{\alpha}$, $\hat{\beta}$) and its covariance estimator $\nabla^2 \chi^2(\cdot, \hat{\beta})$ are asymptotically correct.

Therefore, $\beta = 0$ corresponds to the homogeneity of odds ratios $\theta_k$, the Breslow-Zelen model provides a natural tool to test $\theta_1 = \cdots = \theta_K$. In particular, letting $\chi^2(\cdot, \hat{\beta})$ denote the lower right corner of $\nabla^2 \chi^2(\cdot, \hat{\beta})$, $\chi^2(1-1(\cdot, \hat{\beta}))$ follows a chi-square distribution under the homogeneity assumption and gives a Wald-type test. On the other hand, a score test can be obtained by replacing $\alpha$ by $0$ and $\beta$ by $0$ in (5.1) with a suitable normalization.

6. Applications to Survival Analysis

In this section we apply the results developed earlier to hypothesis testing and parameter estimation in survival analysis, where survival distributions may be discontinuous. We first consider an extension of the proportional hazards model that also covers discontinuity and discuss a normal approximation to the log-rank test statistic. Then we apply the results to a log-rank-type test in survival analysis, which is connected to a group sequential design.

To fix notation, let $T_1, T_2, \ldots, T_n$ denote i.i.d. survival times from the first population with a possibly discontinuous distribution function $F_1$, and $T_{n+1}, \ldots, T_{n+m}$ be i.i.d. survival times from the second population with distribution function $F_2$. Let $\Delta_i, i = 1, 2$ be the corresponding cumulative hazard functions. The usual right censorship is incorporated as we only observe $T_i = T_i \wedge \xi_i, i = 1, \ldots, n_1 + n_2$, where the $\xi_i$ are the censoring times, assumed as usual to be independent of the $T_i$.

In his fundamental paper on the proportional hazards model, Cox (1972) also introduced an extension to accommodate discontinuity by assuming that the odds ratio of the hazard functions to be proportional:

$$\Delta_i(0) = \theta \Delta_i(0) - \Delta_{i+1}(0).$$

(6.1)

Note that when the baseline cumulative hazard function $H(t)$ is continuous, (6.1) effectively reduces to $\Delta_i(0) = \theta \Delta_i(0)$, which is the usual formulation of the proportional hazards model.

Suppose that there are $K$ distinct time points, to be denoted by $t_1 < \cdots < t_k$ at which one or more failures have occurred. This means that for each $t_k$, we can find at least one $i$ with $\Delta_i = t_k$ and $\Delta_i = t_k$. Let $X(t_k) = \#(i = 1 \leq i = n_1 : T_i = t_k$ and $\Delta_i = 1)$,

$$N(t_k) = \#(i = 1 \leq i = n_1 + n_2 : T_i \geq t_k),$$

$$N_{(t_k)} = \#(i = 1 \leq i = n_1 : T_i \geq t_k),$$

$$M_{(t_k)} = \#(i = 1 \leq i = n_1 : T_i = t_k, \Delta_i = 1),$$

$$M(t_k) = n_1 + n_2 - M_{(t_k)}.$$  

(6.2)

Since the underlying distributions may be discontinuous, $X(t_k)$ could take any nonnegative integer value. It is not difficult to show that the quantities
so defined constitute a sequence of 2 x 2 tables as described at the beginning of Section 2.0, satisfying (2.1) with common odds ratio \( \theta \). Therefore, the partial likelihood estimator of \( \theta \) can be obtained by maximizing (4.1) or solving (4.2). The asymptotic properties will remain valid if the variance stability condition in Theorems 3.1 and 4.1 are satisfied. The following remark verifies the condition; the proof will be given in Appendix.

Theorem 6.1. Suppose \( \min(n_1, n_2) \rightarrow \infty \) converges to a positive number. Then, under the extended proportional hazards model assumption (6.1), the variance stability condition is satisfied. More precisely

\[
\sum_{x_1=1}^{K} \sum_{x_2=1}^{K} \frac{V_0(x_1, x_2)}{n_1(x_1)n_2(x_2)} \geq \alpha > 0,
\]

for some positive constant \( \alpha \). Here \( n = n_1 + n_2 \) is the total sample size.

Corollary 6.1. Suppose \( \min(n_1, n_2) \rightarrow \infty \) converges to a positive number. Then the maximum partial likelihood estimator \( \hat{\theta} \), which maximizes (4.1) or solves (4.2) is consistent and asymptotically normal.

Recall that (4.2) is the same as \( \sum_{x_1=1}^{K} (X_1(x_1) - \bar{X}_1(x_1))^2 = 0 \). But \( E_{\theta} X_1(x_1) = \sum_{x_1=1}^{K} 1/(1 + \theta^{-1} x_1) \). So an essential step to compute \( \hat{\theta} \) is to find \( \bar{X}_1 \). First, if \( M_1(x_1) = 1 \), then \( \sum_{x_1=1}^{K} (X_1(x_1) - \bar{X}_1(x_1))^2 = 1/(1 + \theta^{-1} x_1)^2 \). In particular, if the underlying distributions are continuous, then \( M_1(x_1) = 1 \) for all \( x_1 \) and (4.2) becomes

\[
\sum_{x_1=1}^{K} \frac{X_1(x_1) - \frac{1}{1 + \theta^{-1} x_1}}{\frac{1}{1 + \theta^{-1} x_1} / N_1(x_1)} = 0,
\]

which is exactly Cox's partial likelihood estimating equation for the two sample problem and whose asymptotic properties can be established via the elegant martingale theory. Without the continuity assumption, however, the martingale and stochastic integration approaches do not appear to be applicable, yet our Corollary 6.1 still applies.

At the end of Section 2, we proposed a resampling scheme to approximate distribution of \( \hat{\theta} \). The proposal is certainly applicable here. Note that such a resampling scheme is very different from bootstrapping the survival times as one would ordinarily do. It will be interesting to compare the two resampling schemes.

Next, we consider the log-rank test statistic under the current setting. Letting \( \hat{\theta} = 1 \), (4.2) may be used for testing null hypothesis \( H_0 : \theta = 1 \). Under the null hypothesis, \( X_1(x_1) \) follows the central hypergeometric distribution and then \( E(X_1(x_1)|M_1(x_1), N_1(x_1)) = 1 - 2 = M_1(x_1) N_1(x_1) / N(x_1) \) and

\[
\text{Var}(X_1(x_1)|M_1(x_1), N_1(x_1)) = \sum_{x_1=1}^{K} M_1(x_1) N_1(x_1) N(x_1) (N_1(x_1) - 1)^{-} (N(x_1)^{-1})^{-1}.
\]

Hence, a natural test statistic is

\[
U = \sum_{x_1=1}^{K} \frac{X_1(x_1) - \bar{M}_1(x_1) \bar{N}_1(x_1)}{\sqrt{\sum_{x_1=1}^{K} M_1(x_1) N_1(x_1) (N_1(x_1) - 1)^{-} (N(x_1)^{-1})^{-1}}}
\]

From Theorems 3.1 and 6.1, we get the following corollary.

Corollary 6.2. Suppose that \( n_1, n_2 \rightarrow \infty \) converges to a positive number and that the variance stability condition holds. Then \( U \) has asymptotically \( N(0,1) \) distribution under the null hypothesis.

Instead of forming a 2 x 2 table at each failure time point \( x_1 \), one may consider other ways of groupings data. In group sequential design, it is often desirable to conduct the interim analysis at time points so that information accumulated between two consecutive analyses stays constant (Pocock, 1977). Since the information under the proportional hazard is approximately proportional to the number of failures, such a goal may be achieved by setting interim analysis so that the number of failures between any two consecutive interim analyses is equal or close to a prefixed number. It can be shown that under such a design, \( U \), the test statistic defined by (6.3), converges to the standard normal under the null hypotheses.

7. Examples

Example 1. To illustrate the preceding inference methods for the common odds ratio, we shall first consider in Table 1 the following data set, taken from Mietel (1963), about a comparison of the effectiveness of 1.5-hour delayed versus immediately injected Pencillin to protect rabbits against lethal injection with beta-hemolytic streptococcus.

Note that the individual likelihood for the first and last tables are identical to one, whence do not contribute to (4.1). The roots of (2.3) for the other three tables are listed in Table 2. Therefore, solving (4.2) by Newton-Raphson method yields our estimator of the common odds ratio as well as its asymptotic confidence interval.

It can be seen from Table 3 that there is a noticeable difference between the MLE and the M+H estimate, in terms of both the estimators and the confidence intervals. In this special case, our maximum partial likelihood estimator is indeed the conditional maximum likelihood estimator (CMLE), since all the tables are independent. It is suggested in Brenlo
<table>
<thead>
<tr>
<th>Penicillin Level</th>
<th>Response</th>
<th>Delay</th>
<th>Cured</th>
<th>Died</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/8</td>
<td>None</td>
<td>1.5 h</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1/4</td>
<td>None</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5 h</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1/8</td>
<td>None</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5 h</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>None</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5 h</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Penicillin Level</th>
<th>Lambda’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
<td>3.186, 1, 0.314</td>
</tr>
<tr>
<td>1/2</td>
<td>5.552, 1.669, 0.599, 0.180, 0, 0</td>
</tr>
<tr>
<td>1</td>
<td>1, 0, 0, 0, 0, 0</td>
</tr>
</tbody>
</table>

(1981), Hasek (1988), and Sannter and Duffy (1989, section 5.5) that for independent tables, CMLE is better than the M-H estimator in terms of asymptotic variance and efficiency. For this special example, to see which method is better numerically, we shall perform a simulation of 10,000 runs, for estimation of log \( \theta \), with the same margins as specified in the example; and the result is shown in Table 4.

Because of the very small sample size, in simulating the tables we encountered cases violating the equation (4.3), in which both our estimator and the M-H estimator will be undefined, so they will give either \(-\infty \) or \( \infty \).

We deleted all these cases (in particular, the percentage of number of vio-

<table>
<thead>
<tr>
<th>True Value</th>
<th>Method</th>
<th>Sample Mean</th>
<th>Sample Variance</th>
<th>MSE</th>
<th>Coverage Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log ( \theta = 0 )</td>
<td>MLE-MH Method</td>
<td>-0.007</td>
<td>0.729</td>
<td>0.729</td>
<td>0.955</td>
</tr>
<tr>
<td>log ( \theta = 0.5 )</td>
<td>MLE-MH Method</td>
<td>0.529</td>
<td>0.729</td>
<td>0.740</td>
<td>0.966</td>
</tr>
<tr>
<td>log ( \theta = 1 )</td>
<td>MLE-MH Method</td>
<td>0.548</td>
<td>0.800</td>
<td>0.802</td>
<td>0.966</td>
</tr>
<tr>
<td>log ( \theta = -0.5 )</td>
<td>MLE-MH Method</td>
<td>-0.522</td>
<td>0.723</td>
<td>0.723</td>
<td>0.970</td>
</tr>
<tr>
<td>log ( \theta = 2.718 )</td>
<td>MLE-MH Method</td>
<td>0.123</td>
<td>0.693</td>
<td>0.694</td>
<td>0.954</td>
</tr>
<tr>
<td>log ( \theta = -1.0 )</td>
<td>MLE-MH Method</td>
<td>-0.101</td>
<td>0.703</td>
<td>0.704</td>
<td>0.950</td>
</tr>
<tr>
<td>log ( \theta = 0.368 )</td>
<td>MLE-MH Method</td>
<td>-0.101</td>
<td>0.738</td>
<td>0.740</td>
<td>0.950</td>
</tr>
<tr>
<td>log ( \theta = 1.5 )</td>
<td>MLE-MH Method</td>
<td>-0.145</td>
<td>0.581</td>
<td>0.583</td>
<td>0.999</td>
</tr>
<tr>
<td>log ( \theta = 4.82 )</td>
<td>MLE-MH Method</td>
<td>-0.149</td>
<td>0.645</td>
<td>0.645</td>
<td>0.996</td>
</tr>
<tr>
<td>log ( \theta = -0.7 )</td>
<td>MLE-MH Method</td>
<td>-0.173</td>
<td>0.456</td>
<td>0.459</td>
<td>0.976</td>
</tr>
<tr>
<td>log ( \theta = 1.35 )</td>
<td>MLE-MH Method</td>
<td>-0.199</td>
<td>0.514</td>
<td>0.515</td>
<td>0.978</td>
</tr>
<tr>
<td>log ( \theta = 2.0 )</td>
<td>MLE-MH Method</td>
<td>-0.252</td>
<td>0.456</td>
<td>0.462</td>
<td>0.976</td>
</tr>
<tr>
<td>log ( \theta = 1.28 )</td>
<td>MLE-MH Method</td>
<td>-0.293</td>
<td>0.325</td>
<td>0.327</td>
<td>0.939</td>
</tr>
<tr>
<td>log ( \theta = 0.82 )</td>
<td>MLE-MH Method</td>
<td>-0.392</td>
<td>0.293</td>
<td>0.294</td>
<td>0.939</td>
</tr>
</tbody>
</table>

(0.019, 0.819, 0.777, 3.393, 3.203, 10.333, 9.966, 21.965, 21.313, 37.397, 37.377, corresponding to \( \log(\theta) = 0, 0.5, 1, 0, -2.5, \) respectively).
It is evident from Table 1 that the advantage of our approach is that it comes closer to the maximum likelihood estimates than the original estimators, indicating that the bias is less than 1%. Indeed, the former seems to have a smaller mean squared error if the odds ratio is not too big or too small (log [OR] > -2.5).4

Examples 1. The data set of Table 5 is taken from Peto et al. (1985); Example 2. The data set of Table 6 is taken from Rohan et al. (1979). For statistical analysis, it was found that the distribution of the estimates is not normal but is approximately normal with a slight positive skew.

Thus, it has been used by Gehan (1965), Cox (1972), and others. Note: There are many tests for this data set. To analyze the data, we have used a test for two proportions. Because of the constraints, the tables so constructed are dependent. It is also of interest to note that the odds ratio is not significantly different between the two groups. Furthermore, the empirical power functions of the tests are not specified. The only information available is the asymptotic power of the test.

The results are reported in Table 7. Note that except for some assumptions about the covariates, it is impossible to provide exact confidence intervals, even for the conventional approach. Because the empirical power functions of the test are not specified, we have used an approximate standard normal distribution to assess the power of the test. The power is computed by

\[ P = \Phi(z) - \Phi(z - 2.5) \]

where \( \Phi \) is the standard normal cumulative distribution function. The results are reported in Table 7. Note that except for some assumptions about the covariates, it is impossible to provide exact confidence intervals, even for the conventional approach. Because the empirical power functions of the test are not specified, we have used an approximate standard normal distribution to assess the power of the test. The power is computed by

\[ P = \Phi(z) - \Phi(z - 2.5) \]
within Breslow’s confidence interval.

We would also like to point out that, ignoring the dependence between these two by two tables, and formally using the M-H estimator of the common odds ratio, and its variance estimator given in Robins et al. (1986), gives point estimator 5.22 and a 95% confidence interval (2.19, 12.43), which are similar to the ones obtained by the partial likelihood inference and Efron method. However, it remains to be seen whether the Mantel-Haenszel method still gives an asymptotically valid inference, in presence of unknown dependent structures between the tables.

Example 3. To show that our method can be used to handle computation involving large tables, we take the following example from Tuyns et al. (1977) and Breslow and Day (1980). Cases were 200 males with esophageal cancer in a French hospital between January 1972 and April 1974; controls were randomly selected 775 adult males. The table below refer exclusively the role of alcohol for esophageal cancer.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Daily alcohol consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80+ g</td>
</tr>
<tr>
<td>25-34 Case</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>9</td>
</tr>
<tr>
<td>35-44 Case</td>
<td>4</td>
</tr>
<tr>
<td>Control</td>
<td>26</td>
</tr>
<tr>
<td>45-54 Case</td>
<td>25</td>
</tr>
<tr>
<td>Control</td>
<td>29</td>
</tr>
<tr>
<td>55-64 Case</td>
<td>42</td>
</tr>
<tr>
<td>Control</td>
<td>27</td>
</tr>
<tr>
<td>65-74 Case</td>
<td>19</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
</tr>
<tr>
<td>75+ Case</td>
<td>5</td>
</tr>
<tr>
<td>Control</td>
<td>0</td>
</tr>
</tbody>
</table>

Total sample size: 975

As indicated in Breslow and Day (1985, p. 146), standard tests show no evidence of heterogeneity of the odds ratio. Therefore we may assume that \( \hat{\theta} = \cdots = \hat{\theta}_K \). The common odds ratio estimator given by (4.2) is 5.25, along with 95% confidence interval by (4.7) being (3.63, 7.60), compared with M-H estimator 5.16, and the Robins, Breslow, and Greenland (1986) confidence interval (3.56, 7.47). Because of the relatively large sample size, the difference between two methods is not big. The main point here is that it only took less than one second on a Pentium 1.6MHz CPU to get the estimator and confidence interval, even for this case with relatively large sample size.

Appendix: Proofs

Proof of Theorem 4.1. To prove consistency of \( \hat{\theta} \) it suffices to show, in view of (4.2), that for any \( 0 < \epsilon < \hat{\theta} \), there exists a constant \( M(\epsilon) > 0 \) such that

\[
P \left( \inf_{\bar{\theta} \geq \epsilon} \frac{\sum_{k=1}^K Y_k^{(k)}}{\sum_{k=1}^K \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1}} \sum_{k=1}^K Var \bar{\theta}^{-1} X_k^{(k)} \right) > M(\epsilon) \quad \rightarrow \quad 1,
\]

as (3.2) is assumed. In the mean time, by Theorem 3.1 and again (3.2),

\[
\frac{\sum_{k=1}^K X_k^{(k)}}{\sum_{k=1}^K Var \bar{\theta}^{-1} X_k^{(k)}} \rightarrow \frac{\epsilon}{\epsilon - 1}.
\]

Therefore, (A.1) holds if we can show that

\[
\lim_{\epsilon \rightarrow 0} \inf_{\bar{\theta} \geq \epsilon} \frac{\sum_{k=1}^K \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1} - \sum_{k=1}^K \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1}}{\sum_{k=1}^K Var \bar{\theta}^{-1} X_k^{(k)}} \geq \frac{M(\epsilon)}{2}.
\]

Notice that, in the view of the monotonicity of \( (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1} \), the "inf" in (A.2) can only be achieved at \( \bar{\theta} = \hat{\theta} \epsilon \), and also by the mean-value theorem

\[
\sum_{k=1}^K \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1} - \sum_{k=1}^K \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1} = \sum_{k=1}^K \frac{\partial}{\partial \bar{\theta}} \sum_{j=1}^{Y_k^{(k)}} (1 + \bar{\theta}^{-1} A_{1k}^{(k)})^{-1} \bigg|_{\bar{\theta} = \hat{\theta} \epsilon} = \sum_{k=1}^K \frac{2 Y_k^{(k)} \hat{\theta}^{-2} A_{1k}^{(k)}}{(1 + \hat{\theta}^{-1} A_{1k}^{(k)})^2} (\bar{\theta} - \hat{\theta}).
\]
for some \( \delta \), between \( \theta \) and \( \hat{\theta} \). Thus we get

\[
\left( \theta - \delta \right) \frac{\exp \left( -\sum_{k=1}^{n} \frac{X(k)}{k} \right)}{\hat{\theta}^{-1} \lambda(\hat{\theta})} \leq \max(\alpha, \beta) \left( \theta - \delta, \hat{\theta} \right) \frac{\exp \left( -\sum_{k=1}^{n} \frac{X(k)}{k} \right)}{\hat{\theta}^{-1} \lambda(\hat{\theta})} \leq \max(\alpha, \beta) \left( \theta - \delta, \hat{\theta} \right)
\]

where the second inequality follows from the following elementary inequality

\[
\min(\alpha, \beta) \beta(\hat{\theta}) - \gamma \leq \max(\alpha, \beta) \beta(\hat{\theta}) - \gamma
\]

for any \( 0 < \alpha < \infty \) and any \( 0 < \beta < \infty \). Hence (A.2) and (A.1) hold.

To show the asymptotic normality, we observe again by the mean-value theorem

\[
X(k) = \sum_{i=1}^{k} \frac{X(i)}{i} = \frac{X(k)}{k} \sum_{i=1}^{k} \frac{1}{i} = \frac{X(k)}{k} - \frac{1}{k+1}
\]

so that

\[
\frac{X(k)}{k} \sim \frac{1}{k} \frac{1}{k+1}
\]

for some \( \alpha \), between \( \theta \) and \( \hat{\theta} \). Then, along with the fact

\[
\min(\hat{\theta}, \beta, \beta) \frac{\exp \left( -\sum_{k=1}^{n} \frac{X(k)}{k} \right)}{\hat{\theta}^{-1} \lambda(\hat{\theta})} \leq \max(\hat{\theta}, \beta, \beta) \frac{\exp \left( -\sum_{k=1}^{n} \frac{X(k)}{k} \right)}{\hat{\theta}^{-1} \lambda(\hat{\theta})} \leq \max(\hat{\theta}, \beta, \beta) \frac{\exp \left( -\sum_{k=1}^{n} \frac{X(k)}{k} \right)}{\hat{\theta}^{-1} \lambda(\hat{\theta})}
\]

from the inequality (A.3), and the fact \( \hat{\theta} \theta \rightarrow 1 \) in probability, yields

\[
\frac{1}{n} \sum_{k=1}^{n} \frac{X(k)}{k} \sim \frac{1}{n} \hat{\theta} \theta \rightarrow 1
\]

which converges to a nonrandom constant by the law of large numbers.

We next deal with the set \( D \). By definition, the number of points in \( D \), denoted by \( r = r(c) \), is finite. Denote \( D \) the points in \( D \) to be \( \E_{1}, \E_{2}, \ldots, \E_{r} \). Then, as \( n \rightarrow \infty \), the conditional variances of all tables constructed on the times \( t_{1}, \ldots, t_{r} \) will all go to \( 0 \). Therefore, from Kou and Ying (1996), the conditional variances can be approximated by

\[
\frac{1}{n} \sum_{k=1}^{n} \frac{X(k)}{k} \sim \frac{1}{n} \frac{1}{n} \frac{1}{n} \sim \frac{1}{n} \frac{1}{n} \frac{1}{n} \sim \frac{1}{n} \frac{1}{n} \frac{1}{n}
\]

which converges to a nonrandom constant by the law of large numbers.

Finally, from (A.3) we can bound the conditional variance of \( X(k) \) at \( \theta \).
with that $\delta = 1$ to get

$$\sum_{k=\infty}^{\infty} \sum_{\alpha,\beta \in D^2} \nabla \theta^{(k)}(X^{(\alpha)}) \leq \max(\theta, 1) \sum_{k=\infty}^{\infty} \sum_{\alpha,\beta \in D^2} M^{(\alpha)} M^{(\beta)}$$

$$\leq \max(\theta, 1) \sum_{\alpha,\beta \in D^2} M^{(\alpha)}$$

$$\leq \max(\theta, 1) \sum_{\alpha \in D^2} M^{(\alpha)}$$

as $n \to \infty$, where the last inequality follows from the law of large numbers. The above quantity can be made arbitrarily small by letting $\epsilon \to 0$.

Putting these pieces together, we can conclude that $n^{-1} \sum_{k=\infty}^{\infty} \nabla \theta^{(k)}(X^{(\alpha)})$ converges to a constant which is obviously positive. Therefore, the variance stability condition is verified.

Acknowledgements

This research was supported in part by grants from the National Science Foundation and the National Institutes of Health.

References

3. Breslow (1985), Odds ratio estimates when the data are sparse, Biometrika, 62, 73-81.
17. Gart, J. (1976), Point and interval estimation of the common odds ratio in the combination of $n \times 2$ tables, Biometrika, 87, 471-475.
23. Liang, K. Y. and Self, S. F. (1985), Tests for homogeneity of odds ratios when the data are sparse, Biometrika, 72, 354-358.
27. Mathlof (1990), S-plus Manuals, Mathloft: Seattle.
32. Poreck, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, Biometrika, 64, 175-188.
the common odds ratio in several 2 X 2 tables, with discussion. J. of Amer. Stat. Assoc., 90, 1296-1320.
38. Terence, R. E. (1986), On heterogeneity tests based on efficient scores, Biometrika, 72, 91-95.

A NEW TEST OF SYMMETRY ABOUT AN UNKNOWN MEDIAN

WEI WEN MIAO
Macalester College, Department of Mathematics and Computer Science
St. Paul, MN 55105, U.S.A.

A NEW TEST OF SYMMETRY ABOUT AN UNKNOWN MEDIAN

YULIA B. GEL
Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, ON, Canada N2L 3G1

JOSEPH L. GASTWIRTH
Department of Statistics, George Washington University
Washington, D.C. 20052, U.S.A.

Many robust estimators of location, e.g. trimmed means, implicitly assume that the data come from a symmetric distribution. Consequently, it is important to check this assumption with an appropriate statistical test that does not assume a known value of the median or location parameter. This article replaces the mean and standard deviation in the classic Hotelling-Tukey measure of asymmetry by corresponding robust estimators; the median and sample deviation from the median. The asymptotic distribution theory of the test statistic is developed and the new procedure is compared to tests recently proposed by Caball and Maroco (1998) and Miura (1999). Using their approach to approximating the variance of this class of statistics, it is shown that the new test has greater power than the existing tests to detect the asymmetry of skewed contaminated normal data as well as a majority of skewed distributions belonging to the bounded family. The increased power of the new test suggests that the use of robust estimators in goodness of fit type tests deserves further study.

Some key words: Contaminated data; Large sample theory; Mean deviation from the median; Robust estimators; Skewness; Testing symmetry.

1. Introduction
Let $X_1, \ldots, X_n$ be an independent and identically distributed (i.i.d.) sample from an absolutely continuous distribution $F$ with unknown mean $\mu$,.