In "Random Walk, Sequantial Analysis and Related Topics," edited by A. C. Hsiung, Zhiliang Ying and Cun-Hui Zhang, World Scientific, 2006, pp. 171-198.

ANALYSIS OF A SEQUENCE OF DEPENDENT 2 × 2 TABLES

S. G. KOU

Department of IEOR, 312 Mudd Building, Columbia University
500 West 120th Street, New York, NY 10027, U.S.A.

ZHILIANG YING

Department of Statistics, MC 4690, 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, 1996). 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 Breslow-Zelen 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_i and N_i is noncentral hypergeometric

$$P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{M_1 - x}\theta^x}{\sum_{u=L}^{S}\binom{N_1}{u}\binom{N_2}{M_1 - u}\theta^u}, \quad L \le x \le S, \tag{1.1}$$

where $\theta = \{p_1/(1-p_1)\}/\{p_2/(1-p_2)\}$ is the odds ratio parameter, $L = \max(0, M_1 - N_2)$ 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

$$P(X = x) = \binom{N_1}{x}\binom{N_2}{M_1 - x} \Big/ \binom{N}{M_1}, \quad L \le x \le S, \qquad (1.2)$$

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 relationship between occurrence of a disease and exposure to a possible risk factor. Such studies may be summarized into a 2×2 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).

	Diseased	Disease-free	
Exposed	X	$N_1 - X$	N_1
Unexposed	$M_1 - X$	$X + N_2 - M_1$	N_2
	M_1	M_2	N

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 2×2 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 2×2 tables. Such dependent sequences also appear 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 2×2 tables arise from medical follow-up studies.

Due to the nature of follow-up studies, the 2×2 tables thus constructed are typically dependent in such a way that the conditional distribution of the kth table given its margin is noncentral hypergeometric of 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* 2×2 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 2×2 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 are 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 the usual asymptotic properties 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 2×2 tables is due to Mantel and Haenszel (1959). 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.

Third, when the odds ratios are not equal, but dependent on covariates, Breslow (1976) and Zelen (1971) proposed a regression modeling approach. Our method provides an easy way to compute the estimators for such a model, along with their asymptotic confidence intervals. In addition, we are able to justify the asymptotic results even for dependent tables. See Section 5.

Finally, it is common for survival data to have ties, which may be due to the discreteness of the underlying survival distribution, or to grouping of data. In his fundamental paper on the proportional hazards model, Cox (1972) defined an extension to cover possible discontinuity of the survival distribution. The exact partial likelihood in this case appears to be complicated, and various approximations have been proposed (Breslow, 1976, Efron, 1977). The approach proposed in this paper for dependent 2×2 tables is also an effective tool to handle the extension. In particular, it can be used to compute the maximum partial likelihood estimator exactly, not via approximation, and to derive its asymptotic properties; see Section 6.

The paper is organized as follows. Basic decomposition for single table, and tangle sequence is established in the next section. Weighted log-rank test is analyzed in Section 3. Section 4 deals with the problem of maximum partial likelihood estimation, whose asymptotic properties are studied. A useful and slightly more general model proposed by Zelen (1971) and Breslow (1976) is studied in Section 5. Applications to survival data are given in Section 6. Several examples are presented in Section 7 to illustrate the methods. Some technical proofs are given in the Appendix.

2. Preliminary Results

After introducing basic notation, we shall present two results necessary for our discussion. The first is a decomposition of the non-central hypergeometric random variable into a sum of independent, yet not identically distributed, Bernoulli random variables. The second is a "decoupling" result which reduces the study of dependent 2×2 tables to a sequence of "tangent" yet independent sequences.

2.1. *Notation*

Throughout the rest, $\{X^{(k)}, N_i^{(k)}, M_i^{(k)},\ i = 1,2,\ k \geq 1\}$ will be used to describe a sequence of 2×2 tables in which $N_i^{(k)}$ and $M_i^{(k)}$, $i = 1, 2$, are the margins of the kth table and $X^{(k)}$ its upper left corner. Set $N^{(k)} = N_1^{(k)} + N_2^{(k)}$. Note that $N_1^{(k)} + N_2^{(k)} = M_1^{(k)} + M_2^{(k)}$. Associated with the sequence

is a σ-filtration $\{\mathcal{F}_k, k \geq 0\}$ such that $X^{(k)}$ is measurable with respect to $\mathcal{F}_k$ and $N_i^{(k)}$ and $M_i^{(k)}$ are predictable, i.e., measurable with respect to $\mathcal{F}_{k-1}$). The observed data will be K such tables, $\{X^{(k)}, N_i^{(k)}, M_i^{(k)},\ i = 1,2,\ 1 \leq k \leq K\}$. It is assumed that K is a stopping time with respect to $\mathcal{F}_k$ and the conditional distribution of $X^{(k)}$ given $\mathcal{F}_{k-1}$ is noncentral hypergeometric as defined by (1.1), i.e.,

$$P(X^{(k)} = x | \mathcal{F}_{k-1}) = \frac{\binom{N_1^{(k)}}{x}\binom{N_2^{(k)}}{M_1^{(k)}-x}\theta_k^x}{\sum_{u=L^{(k)}}^{S^{(k)}}\binom{N_1^{(k)}}{u}\binom{N_2^{(k)}}{M_1^{(k)}-u}\theta_k^u} \tag{2.1}$$

for $L^{(k)} \leq x \leq S^{(k)}$, where $L^{(k)} = \max(0, M_1^{(k)} - N_2^{(k)})$, $S^{(k)} = \min(N_1^{(k)}, M_1^{(k)})$ and $\theta_k \in (0, \infty)$ is the odds ratio parameter of the kth table. This assumption is satisfied by tables arising from analysis of survival data (Miller, 1981, Chapter 4) as well as medical follow-up studies.

2.2. *A decomposition for a single 2×2 table*

One of the two key elements to derive the main result is the following lemma, which can be found in Kou and Ying (1996).

Lemma 2.1. *Consider a single noncentral hypergeometric random variable X specified by (1.1). Let $\eta_1, \eta_2, \ldots$ be independent uniform $(0,1)$ random variables. Then*

$$X \stackrel{d}{=} \sum_{i=1}^{S} I(\eta_i \leq (1 + \theta^{-1}\lambda_i)^{-1}), \tag{2.2}$$

where "$\stackrel{d}{=}$" denotes equality in distribution, $I(\cdot)$ the indicator function and $-\lambda_1, \ldots, -\lambda_S$ the roots of polynomial

$$\phi(z) = \sum_{u=L}^{S}\binom{N_1}{u}\binom{N_2}{M_1-u}' z^u. \tag{2.3}$$

Polynomial ϕ is, up to a scale constant, the probability generating function of (central) hypergeometric distribution (1.2). A crucial aspect of this lemma is that all roots of ϕ are real, which can be obtained by connecting (2.3) to the classical Jacobi polynomials (Szegö, 1959). Lemma 2.1 is useful not only because it decomposes any noncentral hypergeometric distribution into a sum of independent Bernoulli random variables, but also that all the λ_i are *independent* of the odds ratio θ, as (2.3) does not involve θ. In other

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 kth table, the roots $\lambda_i^{(k)}$ are simple functions of roots of the Jacobi polynomials, which are widely used in mathematics and engineering literature (Kou and Ying, 1996). The roots can be determined easily by using software packages such as *Mathematica* (Wolfram, 1991). Alternatively, one can utilize the associated matrix of which ϕ is the characteristic polynomial. Writing the probability generating function (2.3) as $\phi(z) = \sum_{i=0}^{m} a_i z^i$ $(a_m \neq 0)$, its roots are exactly the eigenvalues of its associate matrix

$$Q = \begin{pmatrix} -\frac{a_{m-1}}{a_m} & -\frac{a_{m-2}}{a_m} & \cdots & -\frac{a_1}{a_m} & -\frac{a_0}{a_m} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

See Johnson and Riess (1982, Section 4.4.2). Again, the eigenvalues can be determined easily by the existing software packages. For example, a relevant command in *Splus* (MathSoft, 1995) is `eigen(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 λ_i and θ:

$$\mathsf{E}_\theta X = \sum_{i=1}^{S} \frac{1}{1+\theta^{-1}\lambda_i}, \quad \mathsf{Var}_\theta(X) = \sum_{i=1}^{S} \frac{\theta^{-1}\lambda_i}{(1+\theta^{-1}\lambda_i)^2}, \tag{2.4}$$

where the subscript θ in E_θ and Var_θ indicates that the expectation and variance are taken with θ 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, $\{Y_k, k \geq 1\}$ and $\{Z_k, k \geq 1\}$, are said to be tangent with respect to $\{\mathcal{H}_k\}$ if, for each k, Y_k and Z_k are $\mathcal{H}_k$ measurable and have the same conditional distribution given $\mathcal{H}_{k-1}$, i.e.,

$$Y_k|\mathcal{H}_{k-1} \stackrel{d}{=} Z_k|\mathcal{H}_{k-1}.$$

Definition 2.2. A sequence of random variables $\{Y_k\}$ adapted to $\{\mathcal{H}_k\}$ (i.e. Y_k is measurable with respect to $\mathcal{H}_k$) is said to satisfy condition (CI) with respect to $\mathcal{T}$, if there exists a σ-field $\mathcal{T} \subset \bigvee_{k\geq 1} \mathcal{H}_k$, where $\bigvee_{k\geq 1} \mathcal{H}_k$ denotes the minimum σ-field spanned by $\{\mathcal{H}_l, l \geq 1\}$, such that $Y_k|\mathcal{H}_{k-1} \stackrel{d}{=} Y_k|\mathcal{T}$ for all $k \geq 1$ and that Y_k are conditionally mutually independent given $\mathcal{T}$.

Lemma 2.2. *(See Theorem 5.8.3 of Kwapień and Woyczyński (1992) and also de la Peña and Giné (1999)). For each $n \geq 1$, let $\{Y_{nk}, k \geq 1\}$ be a tangent sequence of $\{Z_{nk}, k \geq 1\}$ with respect to σ-filtration $\{\mathcal{H}_{nk}, k \geq 1\}$. Suppose, for each $n \geq 1$, the sequence $\{Y_{nk}\}$ also satisfies condition (CI) with respect to $\{\mathcal{T}_n\}$. If the conditional distribution of $\sum_k Y_{nk}$ given $\mathcal{T}_n$ converges to a non-random probability distribution μ, then $\sum_k Z_{nk}$ also converges in distribution to μ.*

3. Weighted Log-Rank Test Statistic

Consider now a normalized sum of weighted 2×2 tables

$$U = \sum_{k=1}^{K} w_k \left(X^{(k)} - \mathsf{E}_{\theta_k}^{(k-1)} \left(X^{(k)} \right) \right) \Big/ \left(\sum_{k=1}^{K} w_k^2 \mathsf{Var}_{\theta_k}^{(k-1)} (X^{(k)}) \right)^{1/2}, \tag{3.1}$$

where w_k are $\mathcal{F}_{k-1}$-measurable weights, and $\mathsf{E}_{\theta_k}^{(k-1)}$, $\mathsf{Var}_{\theta_k}^{(k-1)}$ denote conditional expectation and variance given $\mathcal{F}_{k-1}$ with θ_k being the odds ratio of the kth table. An important special case is $\theta_k \equiv 1$ for all k. In such a case, $\mathsf{E}^{(k-1)}(X^{(k)}) = M_1^{(k)} N_1^{(k)}/N^{(k)}$, $\mathsf{Var}^{(k-1)}(X^{(k)}) = M_1^{(k)} M_2^{(k)} N_1^{(k)} N_2^{(k)}/\{(N^{(k)}-1)(N^{(k)})^2\}$ and U is known as the normalized weighed log-rank test statistic for testing null hypothesis $\theta_k = 1$ (Tarone and Ware, 1977). Because the exact distribution for such a statistic is practically impossible to derive, and furthermore the dependency structure among tables may be complicated, the asymptotic approximation becomes essential for inference.

To apply the decoupling method, we need to construct a tangent sequence of $\{X^{(k)}\}$ that also satisfies condition (CI). Let $\{\eta_j, j \geq 1\}$ be a sequence of independent random variables with the uniform distribution on $(0,1)$. They are chosen, in particular, to be independent of filtration $\{\mathcal{F}_k, k \geq 0\}$, thus also of the original sequence of 2×2 tables. For each

k and each $i = 1, \ldots, S^{(k)}$, let $\eta_{k,i} = I(k \le K)\eta_{S^{(1)}+\cdots+S^{(k-1)}+i}$, and let $\mathcal{G}_k$ denote the σ-field generated by $\mathcal{F}_k$ and $\{\eta_{l,j}, j = 1, \ldots, S^{(l)}, l = 1, \ldots, k\}$. Then by Lemma 2.1, there exist, for each $1 \le k \le K$, nonnegative numbers $\lambda_1^{(k)}, \ldots, \lambda_{S^{(k)}}^{(k)}$, which depend *only* on $N_1^{(k)}, N_2^{(k)}, M_1^{(k)}, M_2^{(k)}$, such that conditional on these margins,

$$\sum_{i=1}^{S^{(k)}} I(\eta_{k,i} \le (1 + \theta_k^{-1}\lambda_i^{(k)})^{-1})$$

has the same hypergeometric distribution as that of $X^{(k)}$. Define

$$X_k^* = w_k[X^{(k)} - \mathsf{E}_{\theta_k}^{(k-1)}(X^{(k)})]I(K \ge k),$$

$$W_k^* = w_k \sum_{i=1}^{S^{(k)}} \left[I(\eta_{k,i} \le (1 + \theta_k^{-1}\lambda_i^{(k)})^{-1}) - (1 + \theta_k^{-1}\lambda_i^{(k)})^{-1}\right] I(K \ge k).$$

Then, recalling the expression of the expectation in (2.4) and $I(K \ge k) \in \mathcal{F}_{k-1}$, we see that W_k^* and X_k^* have the same conditional distribution given $\mathcal{G}_{k-1}$. We therefore have the following conclusion.

Lemma 3.1. *The sequences of random variables $\{X_k^*\}$ and $\{W_k^*\}$ are tangent with respect to $\{\mathcal{G}_k\}$.*

Next let $\mathcal{T} = \bigvee_{k\ge1} \mathcal{F}_k$. Then $\mathcal{T} \subset \bigvee_{k\ge1} \mathcal{G}_k$. Clearly by examining the probability generating functions, we know that $W_k^*|\mathcal{G}_{k-1} \stackrel{d}{=} W_k^*|\mathcal{T}$. Moreover, conditional on $\mathcal{T}$, the $\lambda_i^{(k)}$ are fixed and the $\eta_{k,j}$ are independent, whence W_k^* are independent. So we have the following lemma.

Lemma 3.2. *The sequence $\{W_k^*\}$ satisfies condition (CI) with, using notation in Definitions 2.1 and 2.2, $\mathcal{H}_k = \mathcal{G}_k$ and $\mathcal{T} = \bigvee_{k\ge1} \mathcal{F}_k$.*

In view of the preceding construction, we see that the weighted sum of the X_k can be approximated by a sum of weighted independent Bernoulli random variables, whose asymptotic distribution is easy to characterize and which is also quite simple to simulate. The asymptotic results to be presented below is interpreted in the sense of double arrays: there is an additional index n going to ∞ and all the quantities $(X^{(k)}, M_i^{(k)}, N_i^{(k)}, i = 1, 2, w_k, \mathcal{F}_k), k = 1, \ldots, K$, depend on n.

Theorem 3.1. *Suppose that there exists a nonrandom sequence $q_n \to \infty$ such that*

$$\frac{\sum_{k=1}^K w_k^2 \mathsf{Var}_{\theta_k}^{(k-1)}(X^{(k)})}{q_n} \xrightarrow{p} 1 \qquad (3.2)$$

and that $\max_{1\le k\le K} w_k^2/q_n \to 0$ in probability. Then U defined by (3.1) converges in distribution to $N(0,1)$.

Proof. Let $\tilde{X}_k^* = X_k^*/\sqrt{q_n}$ and $\tilde{W}_k^* = W_k^*/\sqrt{q_n}$. By Lemma 3.1, $\{\tilde{X}_k^*\}$ and $\{\tilde{W}_k^*\}$ are tangent with respect to $\{\mathcal{G}_k\}$; by Lemma 3.2, $\{\tilde{W}_k^*\}$ satisfies condition (CI) with $\mathcal{T} = \bigvee_{k\ge1} \mathcal{F}_k$. Thus by Lemma 2.2, $\sum_{k\ge1} \tilde{X}_k^*$ converges to $N(0,1)$, provided we can show that $\sum_{k\ge1} \tilde{W}_k^*|\mathcal{T}$ converges to standard normal distribution. The latter is obvious because, conditional on $\mathcal{T}$, $\sum_{k\ge1} W_k^* = \sum_{j=1}^{S^{(k)}} w_j[I(\eta_{k,j} \le (1+\theta^{-1}\lambda_j^{(k)})^{-1}) - (1+\theta^{-1}\lambda_j^{(k)})^{-1}]$ is a sum of independent random variables and, by (3.2), $\mathsf{Var}(\sum_{k\ge1} W_k^*|\mathcal{T})/q_n \xrightarrow{p} 1$ which, in conjunction with the assumption $\max_j w_j^2/q_n \to 0$, implies the Lindeberg condition. In view of of (3.2), U also converges to $N(0,1)$, via Slutsky's Theorem. □

It seems that the variance stability by q_n in the theorem is indispensable even in some simplest cases. To give an example, consider a sequence of independent 2×2 tables with common odds ratio $\theta = 1$, and the margins $M_1^{(k)} = M_2^{(k)} = N_1^{(k)} = N_2^{(k)} = 1$. Let K be the first time that $\sum_{k=1}^K (X^{(k)} - 1/2) \ge n$. Then $\sum_{k=1}^K \mathsf{Var}_\theta^{(k-1)}(X^{(k)}) = K/4 \to \infty$ as $n \to \infty$. However $U = (\sum_{k=1}^K (X^{(k)} - 1/2))(K/4)^{-1/2} \ge 0$ and therefore does not converge to $N(0,1)$.

It may be revealing to compare Theorem 3.1 with some known results in special cases. If we consider a special case that K is nonrandom and does not increase to ∞, and that the K tables are independent, then we can apply Lemma 2.1 K times to express U (in distribution) as a weighted sum of independent Bernoulli random variables. Thus U converges in distribution to $N(0,1)$ via the Lindeberg central limit theorem. In another extreme case in which $M_1^{(k)} \equiv 1$, implying $X^{(k)}$ is either 0 or 1, the Lindeberg-type condition for U becomes trivial. It follows from a martingale central limit theorem as given in Pollard (1984, p.171) that U converges in distribution to $N(0,1)$ provided that $K \to \infty$ in probability and an additional variance stability condition holds. In survival analysis, the last situation arises when one constructs a two-sample log-rank test under the assumption that the underlying survival distribution is continuous (therefore no ties). See Miller (1981) and Fleming and Harrington (1991).

In general, for some k, $X^{(k)}$ may have large conditional variances, while for others $X^{(k)}$ may be bounded (thus having relatively small variances). For cases of the first kind, each normalized $X^{(k)}$ should already be closed to normal; and for the latter ones, the martingale central limit theorem suggests intuitively that their normalized sum should also be close to nor-

mal. 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., $\theta_k = \theta$ (Mantel and Haenszel, 1959; Breslow, 1996). Therefore, an important statistical problem is to estimate the common odds ratio θ and that will be the concern of this section.

4.1. *The estimator*

The conditional probability mass function of the kth table given its margins and the $k-1$ preceding tables can be written as

$$g_k(M_1^{(k)}, M_2^{(k)}, N_1^{(k)}, N_2^{(k)}) \frac{\binom{N_1^{(k)}}{X^{(k)}}\binom{N_2^{(k)}}{M_1^{(k)}-X^{(k)}}\theta^{X^{(k)}}}{\sum_{u=L^{(k)}}^{S^{(k)}}\binom{N_1^{(k)}}{u}\binom{N_2^{(k)}}{M_1^{(k)}-u}\theta^u},$$

where g_k is the conditional probability mass of $(M_1^{(k)}, M_2^{(k)}, N_1^{(k)}, N_2^{(k)})$ given the $k-1$ preceding tables. Following Cox (1975), we ignore the g's and obtain the following partial likelihood

$$\prod_{k=1}^{K} \frac{\binom{N_1^{(k)}}{X^{(k)}}\binom{N_2^{(k)}}{M_1^{(k)}-X^{(k)}}\theta^{X^{(k)}}}{\sum_{u=(k)}^{S^{(k)}}\binom{N_1^{(k)}}{u}\binom{N_2^{(k)}}{M_1^{(k)}-u}\theta^u}. \tag{4.1}$$

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 θ equal to 0, we see that the maximum partial likelihood estimator (MPLE), denoted by $\hat{\theta}$, satisfies the following equation

$$\sum_{k=1}^{K}\left[X^{(k)} - E_\theta^{(k-1)}X^{(k)}\right] = 0,$$

which, through (2.4), is the same as

$$\sum_{k=1}^{K} X^{(k)} - \sum_{k=1}^{K}\sum_{i=1}^{S^{(k)}} \frac{1}{1+\theta^{-1}\lambda_i^{(k)}} = 0. \tag{4.2}$$

The simple form of (4.2) allows the following 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*

$$\sum_{k=1}^{K} L^{(k)} < \sum_{k=1}^{K} X^{(k)} < \sum_{k=1}^{K} S^{(k)}. \tag{4.3}$$

Since $L^{(k)} \le X^{(k)} \le S^{(k)}$ for all k, it follows that (4.3) is equivalent to that there exist k and k' such that $L^{(k)} < X^{(k)}$ and $X^{(k')} < S^{(k')}$.

Proof. We first show that (4.3) is sufficient. It is obvious that, on $(0,\infty)$, $(1+\theta^{-1}\lambda_i^{(k)})^{-1}$ is strictly increasing in θ, if $\lambda_i^{(k)} > 0$. Therefore, $\sum_{k=1}^{K}\sum_{i=1}^{S^{(k)}}\{1+\theta^{-1}\lambda_i^{(k)}\}^{-1}$ goes to $\sum_{k=1}^{K} L^{(k)}$ as $\theta \to 0$, and goes to $\sum_{k=1}^{K} S^{(k)}$ as $\theta \to \infty$, because for each k there are exactly $L^{(k)}$ zero among $\{\lambda_i^{(k)},\ 1 \le i \le S^{(k)}\}$. We have then that, for any $X \in (\sum_{k=1}^{K} L^{(k)}, \sum_{k=1}^{K} S^{(k)})$, there exists a unique $\hat{\theta}$ such that (4.2) holds. The same reason also leads to the conclusion that no such $\hat{\theta}$ exists for $X = \sum_{k=1}^{K} L^{(k)}$ or $X = \sum_{k=1}^{K} S^{(k)}$, whence the necessity is obtained as well. □

It is not difficult to see that the partial likelihood estimating 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 $\lambda_i^{(k)}$ 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 (3.2) is satisfied with $w_k \equiv 1$. Then $\hat{\theta}$ is consistent and asymptotically normal. More precisely, we have $\hat{\theta} \to \theta$ in probability and*

$$\theta^{-1}\left(\sum_{k=1}^{K}\mathrm{Var}_\theta^{(k-1)}(X^{(k)})\right)^{1/2}(\hat{\theta}-\theta) \to_{\mathcal{L}} N(0,1), \tag{4.4}$$

$$\left(\sum_{k=1}^{K}\mathrm{Var}_\theta^{(k-1)}(X^{(k)})\right)^{1/2}(\log\hat{\theta}-\log\theta) \to_{\mathcal{L}} N(0,1). \tag{4.5}$$

Furthermore, the normalizing factor, $\left(\sum_{k=1}^{K} \mathrm{Var}_{\theta}^{(k-1)}(X^{(k)})\right)$, *may be replaced by its empirical counterpart* $\widehat{\mathrm{Var}}$, *where*

$$\widehat{\mathrm{Var}} = \sum_{k=1}^{K} \mathrm{Var}_{\hat{\theta}}^{(k-1)}(X^{(k)}) = \sum_{k=1}^{K} \sum_{i=1}^{S^{(k)}} \frac{\hat{\theta}^{-1}\lambda_i^{(k)}}{(1+\hat{\theta}^{-1}\lambda_i^{(k)})^2}, \tag{4.6}$$

and (4.4) and (4.5) still hold with the replacement.

Because of skewness (θ is positive), it is often more desirable to construct asymptotically valid confidence intervals using (4.5) rather than (4.4). In particular, an asymptotic $100(1-\alpha)\%$ confidence interval for $\log\theta$ is

$$\log\hat{\theta} \pm z_{\alpha/2}(\widehat{\mathrm{Var}})^{-1/2}, \tag{4.7}$$

which can be exponentiated to get the corresponding interval for θ

$$\hat{\theta}\exp\{\pm z_{\alpha/2}(\widehat{\mathrm{Var}})^{-1/2}\}.$$

Using (4.2) and (4.5) we can also get a simple and asymptotically valid resampling method to approximate the distribution of $\hat{\theta}$. Specifically, we first calculate $\lambda_i^{(k)}$ and $\hat{\theta}$. We generate uniform(0,1) random variables $\eta_{k,i}$. With $\lambda_i^{(k)}$, $\hat{\theta}$ and $\eta_{k,i}$, we get simulated data via

$$\bar{X}^{(k)} = \sum_{i=1}^{S^{(k)}} I(\eta_{k,i} \le (1+\hat{\theta}_k^{-1}\lambda_i^{(k)})^{-1}).$$

We then compute $\tilde{\theta}$ from the simulated data $\bar{X}^{(k)}$ by solving

$$\sum_{k=1}^{K}\left[\bar{X}^{(k)} - \sum_{i=1}^{S^{(k)}} \frac{1}{1+\tilde{\theta}^{-1}\lambda_i^{(k)}}\right] = 0.$$

It is not difficult to justify that the conditional distribution of $\tilde{\theta}-\hat{\theta}$ does approximate the unconditional distribution of $\hat{\theta}-\theta$. So we can repeatedly generate $\tilde{\theta}$ to obtain empirically the (conditional) distribution of $\tilde{\theta}-\hat{\theta}$, which can then be used for inference of θ.

4.2. *Comparison with other estimators*

Various other methods have been proposed for the inference of the common odds ratio parameter. Woolf (1955) suggested to estimate logarithm of the common odds ratio by a weighted sum of logarithms of the empirical odds ratios from different tables. Validity of his method requires that sizes of all tables be reasonably large. Inconsistency could arise if it is applied to a large number of sparse tables (Breslow, 1981).

The most widely used method is, perhaps, due to Mantel and Haenszel (1959). Let $X_{ij}^{(k)}$, $1 \le i,j \le 2$ be the four cell counts in the kth tables, i.e., $X_{11}^{(k)} = X^{(k)}$, $X_{12}^{(k)} = N_1^{(k)} - X^{(k)}$, etc. The Mantel-Haenszel (M-H) estimator is defined by

$$\hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} X_{11}^{(k)} X_{22}^{(k)}/N^{(k)}}{\sum_{k=1}^{K} X_{12}^{(k)} X_{21}^{(k)}/N^{(k)}}. \tag{4.8}$$

It is known that the estimator is consistent in the case of either a small number of large tables or a large number of sparse tables and is robust against indivisual zero cell entries. The estimator is also obviously easy to compute.

Comparing with the M-H estimator, the computation of the MPLE is more involved. However, as we discussed before, with the modern computing capability, finding MPLE $\hat{\theta}$ has become manageable as the roots λ's can be found easily.

There are two main advantages of the MPLE. First, as Theorem 3.1 shows, the MPLE is consistent for both a finite number of large tables and a large number of sparse tables. In addition, it is also valid for a combination of large and small tables. In contrast, it has been no theoretic justification for the M-H estimator in the case of mixture of large and small tables. Secondly, the variance estimation for $\hat{\theta}_{MH}$ could be quite complicated (Robins et al., 1986; Phillips and Holland, 1987), whereas the variance estimator for the MPLE is given in (4.7).

It is worth mentioning that the MPLE is also robust against individual zero cell entries. In fact, the following theorem shows that the MPLE exists if and only if the M-H exists.

Theorem 4.2. *A necessary and sufficient condition for the existence and uniqueness of MPLE* $\hat{\theta}$ *is both* $\sum_k X_{11}^{(k)} X_{22}^{(k)}$ *and* $\sum_k X_{12}^{(k)} X_{21}^{(k)}$ *are not zero. Therefore, MPLE is well defined if and only if* $\hat{\theta}_{MH}$ *is well defined.*

Proof. By Proposition 3.1 and Remark 3.1, it suffices to show that, for any k and k', $L^{(k)} < X^{(k)}$ and $X^{(k')}$ if and only if $X_{11}^{(k)}X_{22}^{(k)} > 0$ and $X_{12}^{(k')}X_{21}^{(k')} > 0$. But it is straightforward to see from the 2×2 table that $X_{11}^{(k)}X_{22}^{(k)} > 0$ is equivalent to $L^{(k)} < X^{(k)}$ and $X_{12}^{(k')}X_{21}^{(k')} > 0$ is equivalent to $X^{(k')} < S^{(k')}$. So the theorem follows. □

5. Analysis of the Breslow-Zelen Model

The homogeneity assumption about odds ratio parameters θ_k of the K tables may be violated. To accommodate the possible nonhomogeneity, a useful regression model was proposed and studies by Zelen (1971) and Breslow (1976). Their model assumes relationship

$$\log(\theta_k) = \alpha + \beta' \mathbf{z_k}, \qquad 1 \le k \le K,$$

where $\mathbf{z_k}$ is a vector of covariates associated with the kth table, and α and β the intercept and the regression parameters. Typically, the $\mathbf{z_k}$ include the stratification variable used to obtain the sequence of tables. See Breslow and Cologne(1986). Note that without including the $\mathbf{z_k}$, 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 θ replaced by $\exp(\alpha + \beta' z_k)$. Differentiating with respect to α and β, we arrive at the following estimating equation, which is analogous to (4.2),

$$\sum_{k=1}^{K} \left(X^{(k)} - \sum_{i=1}^{S^{(k)}} \frac{1}{1 + \lambda_i^{(k)} / \exp(\alpha + \beta' \mathbf{z_k})} \right) \begin{pmatrix} 1 \\ \mathbf{z_k} \end{pmatrix} = 0. \tag{5.1}$$

The $\lambda_i^{(k)}$ are the same as those in (4.2) and thus depend neither on the parameter value 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 MPLE $(\hat{\alpha}, \hat{\beta})$ of (α, β). The asymptotic variance-covariance matrix can be estimated quiet easily by $\mathbf{I}^{-1}(\hat{\alpha}, \hat{\beta})$, where

$$\mathbf{I}(\alpha, \beta) = \sum_{k=1}^{K} \begin{pmatrix} 1 \\ \mathbf{z_k} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z_k} \end{pmatrix}' \sum_{i=1}^{S^{(k)}} \frac{\lambda_i^{(k)} / \exp(\alpha + \beta' \mathbf{z_k})}{[1 + \lambda_i^{(k)} / \exp(\alpha + \beta' \mathbf{z_k})]^2}.$$

Suppose that there exists a nonrandom sequence q_n such that $q_n^{-1} \mathbf{I}(\alpha, \beta)$ 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 $\mathbf{I}^{-1}(\hat{\alpha}, \hat{\beta})$ are asymptotically correct.

Because $\beta = 0$ corresponds to the homogeneity of odds ratios θ_k, the Breslow-Zelen model provides a natural tool to test $\theta_1 = \cdots = \theta_K$. In particular, letting $(\mathbf{I}^{-1}(\alpha, \beta))_{22}$ denote the lower right corner of $\mathbf{I}^{-1}(\alpha, \beta)$,

$\hat{\beta}' (\mathbf{I}^{-1}(\hat{\alpha}, \hat{\beta}))_{22}^{-1} \hat{\beta}$ follows a χ^2 distribution under the homogeneity assumption and gives a Wald-type test. On the other hand, a score test can be obtained by replacing α by $\hat{\alpha}$ and β 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 problems 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 $T_1, T_2, \ldots, T_{n_1}$ denote i.i.d. survival times from the first population with a possibly discontinuous distribution function F_1 and $T_{n_1+1}, \ldots, T_{n_1+n_2}$ be i.i.d. survival times from the second population with distribution function F_2. Let Λ_i, $i = 1, 2$ be the corresponding cumulative hazard functions. The usual right censorship is incorporated as we only observe $\tilde{T}_i = T_i \wedge C_i$, and $\Delta_i = I(T_i \le C_i)$, $i = 1, \ldots, n_1 + n_2$, where the C_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

$$\frac{\mathrm{d}\Lambda_1(t)}{1 - \mathrm{d}\Lambda_1(t)} = \theta \frac{\mathrm{d}\Lambda_2(t)}{1 - \mathrm{d}\Lambda_2(t)}. \tag{6.1}$$

Note that when the baseline cumulative hazard function Λ_2 is continuous, (6.1) effectively reduces to $\mathrm{d}\Lambda_1(t) = \theta \mathrm{d}\Lambda_2(t)$, which is the usual formulation of the proportional hazards model.

Suppose that there are K distinct time points, to be denoted by $\tau_1 < \cdots < \tau_k$, at which one or more failures have occurred. This means that for each τ_k, we can find at least one i with $\Delta_i = 1$ and $\tilde{T}_i = \tau_k$. Let

$$\begin{aligned} X^{(k)} &= \#\{i \le n_1 : \tilde{T}_i = \tau_k \text{ and } \Delta_i = 1\}, \\ N^{(k)} &= \#\{i \le n_1 + n_2 : \tilde{T}_i \ge \tau_k\}, \\ N_1^{(k)} &= \#\{i \le n_1 : \tilde{T}_i \ge \tau_k\}, \quad N_2^{(k)} = N^{(k)} - N_1^{(k)}, \\ M_1^{(k)} &= \#\{i \le n_1 + n_2 : \tilde{T}_i = \tau_k, \ \Delta_i = 1\}, \quad M_2^{(k)} = N^{(k)} - M_1^{(k)}. \end{aligned} \tag{6.2}$$

Since the underlying distributions may be discontinuous, $X^{(k)}$ could take any nonnegative integer value. It is not difficult to show that the quantities

so defined constitute a sequence of 2×2 tables as described at the beginning of Section 2, satisfying (2.1) with common odds ratio θ. Therefore, the partial likelihood estimator of θ 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 result verifies the condition; the proof will be given in Appendix.

Theorem 6.1. *Suppose* $\min\{n_1, n_2\}/n$ *converges to a positive number. Then, under the extended proportional hazards model assumption (6.1), the variance stability condition is satisfied. More precisely*

$$\frac{\sum_{k=1}^{K} \mathrm{Var}_{\theta}^{(k-1)}(X^{(k)})}{n} \xrightarrow{p} \alpha > 0,$$

for some positive constant α. *Here* $n = n_1 + n_2$ *is the total sample size.*

Corollary 6.1. *Suppose* $\min\{n_1, n_2\}/n$ *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 $\sum_{k=1}^{K}(X^{(k)} - \mathrm{E}_{\theta}^{(k-1)} X^{(k)}) = 0$. But $\mathrm{E}_{\theta} X^{(k)} = \sum_i 1/(1 + \theta^{-1}\lambda_i^{(k)})$. So an essential step to compute $\hat{\theta}$ is to find $\lambda_i^{(k)}$ first. However, if $M_1^{(k)} = 1$, i.e. only one failure occurs at τ_k, then $\mathrm{E}_{\theta}^{(k-1)} X^{(k)} = 1/(1 + \theta^{-1} N_2^{(k)}/N_1^{(k)})$. In particular, if the underlying distributions are continuous, then $M_1^{(k)} = 1$ for all k and (4.2) becomes

$$\sum_{k=1}^{K} \left(X^{(k)} - \frac{1}{1 + \theta^{-1} N_2^{(k)}/N_1^{(k)}} \right) = 0,$$

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 4.1, 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 $\theta = 1$, (4.2) may be used for testing null hypothesis $F_1 = F_2$. Under the null hypothesis, $X^{(k)}$ follows the central hypergeometric distribution and thus $\mathrm{E}(X^{(k)} | M_i^{(k)}, N_i^{(k)},\ i = 1, 2) = M_1^{(k)} N_1^{(k)}/N^{(k)}$ and

$\mathrm{Var}(X^{(k)} | M_i^{(k)}, N_i^{(k)}) = \sum_{k=1}^{K} M_1^{(k)} M_2^{(k)} N_1^{(k)} N_2^{(k)} (N^{(k)} - 1)^{-1} (N^{(k)})^{-1}$. Hence, a natural test statistic is

$$U = \frac{\sum_{k=1}^{K} \left(X^{(k)} - M_1^{(k)} N_1^{(k)}/N^{(k)} \right)}{\left[\sum_{k=1}^{K} M_1^{(k)} M_2^{(k)} N_1^{(k)} N_2^{(k)} \left(N^{(k)} - 1 \right)^{-1} \left(N^{(k)} \right)^{-2} \right]^{1/2}}. \tag{6.3}$$

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

Corollary 6.2. *Suppose that* n_1/n *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 2×2 table at each failure time point τ_k, one may consider other ways of grouping 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 are 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 hypothesis.

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 Mantel (1963), about a comparison of the effectiveness of 1.5-hour-delayed versus immediately injected Penicillin to protect rabbits against lethal injection with β-hemolytic streptococci.

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 MPLE 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 Breslow

Penicillin Level	Delay	Response Cured	Response Died
1/8	None	0	6
	1.5 h	0	5
1/4	None	3	3
	1.5 h	0	6
1/2	None	6	0
	1.5 h	2	4
1	None	5	1
	1.5 h	6	0
4	None	2	0
	1.5 h	5	0

Penicillin Level	Lambda's
1/4	3.186, 1, 0.314
1/2	5.552, 1.669, 0.599, 0.180, 0, 0
1	1, 0, 0, 0, 0, 0

Method	Estimator of θ	95% C. I. for θ
MPLE	10.36	(1.13, 94.77)
M-H method	7	(1.03, 47.73)

(1981), Hauck (1988), and Santner 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, as they will give either $-\infty$ or ∞. We deleted all these cases (in particular, the percentage of number of violation for the simulations listed in Table 4 are 0.19%, 0.81%, 0.77%, 3.29%, 3.20%, 10.33%, 9.90%, 21.96%, 21.31%, 37.39%, 37.37%, corresponding to $\log(\theta) = 0$, 0.5, -0.5, 1.0, ..., -2.5, respectively).

Ture Value	Method	Sample Mean	Sample Variance	MSE	Coverage Prob.
$\log\theta = 0$	MPLE Method	-0.007	0.729	0.729	0.965
$(\theta = 1)$	M-H method	-0.008	0.789	0.789	0.965
$\log\theta = 0.5$	MPLE Method	0.529	0.739	0.740	0.966
$(\theta = 1.649)$	M-H method	0.548	0.800	0.802	0.966
$\log\theta = -0.5$	MPLE Method	-0.522	0.723	0.723	0.970
$(\theta = 0.607)$	M-H method	-0.541	0.783	0.785	0.970
$\log\theta = 1.0$	MPLE Method	1.023	0.693	0.694	0.954
$(\theta = 2.718)$	M-H method	1.062	0.760	0.764	0.954
$\log\theta = -1.0$	MPLE Method	-1.012	0.689	0.689	0.950
$(\theta = 0.368)$	M-H method	-1.051	0.758	0.761	0.950
$\log\theta = 1.5$	MPLE Method	1.426	0.587	0.593	0.988
$(\theta = 4.82)$	M-H method	1.479	0.649	0.650	0.985
$\log\theta = -1.5$	MPLE Method	-1.445	0.581	0.583	0.990
$(\theta = 0.223)$	M-H method	-1.499	0.645	0.645	0.986
$\log\theta = 2.0$	MPLE Method	1.725	0.455	0.531	0.979
$(\theta = 7.389)$	M-H method	1.792	0.514	0.557	0.978
$\log\theta = -2.0$	MPLE Method	-1.730	0.456	0.529	0.976
$(\theta = 0.135)$	M-H method	-1.799	0.518	0.558	0.976
$\log\theta = 2.5$	MPLE Method	1.950	0.325	0.627	0.939
$(\theta = 12.182)$	M-H method	2.025	0.378	0.604	0.939
$\log\theta = -2.5$	MPLE Method	-1.942	0.329	0.641	0.932
$(\theta = 0.082)$	M-H method	-2.020	0.385	0.616	0.932

Drug 6-MP	6+, 6, 6, 6, 7, 9+, 10+, 10, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+
Control	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

It is evident from Table 4 that an advantage of our approach is that it has a smaller variance than that of the M-H estimator, uniformly for the cases shown in Table 4. Indeed, the former seems to have smaller mean squared error if the odds ratio is not too big or too small ($|\log(\theta)| < 2.5$); otherwise both estimators are substantially biased (it might well be these cases are too extreme in view of the small sample size, as in such cases about 37% of the simulations lead to undefined estimators). Because of symmetry, the results are similar for positive and negative $\log(\theta)$. It also appears that both confidence intervals have roughly same coverage probabilities.

Example 2. The data set of Table 5 is taken from Freireich et al. (1963), containing times (weeks) of remission of leukemia patients from two groups, drug 6-MP group and control group. The plus signs denote censored values. It has been used by Gehan (1965), Cox (1972) among others.

Note that there are many ties in this data set. To analyze the data, we construct a 2×2 table at each failure time, resulting in 17 tables as shown in Table 6.

Because of the censorship, the tables so constructed are *dependent*, and, furthermore, the dependent structures are not specified; the only information we have is that for each table conditioning on the margins (N_1, N_2, M_1, M_2), the first cell (X) is hypergeometrically distributed. Let $\Lambda_1(t)$ and $\Lambda_0(t)$ be cumulative hazard rates for the control and 6-MP groups, respectively. Assume that the hazard rates are proportional for the two groups

$$\frac{d\Lambda_1(t)}{1-d\Lambda_1(t)} = e^{\beta}\frac{d\Lambda_0(t)}{1-d\Lambda_0(t)},$$

where $e^{\beta} = \theta$ is the common odds ratio for the 17 tables.

We can use the maximum partial likelihood estimator via (4.2) and (4.7). The results are reported in Table 7. Note that without further assumptions about censorship, it is impossible to provide an exact confidence interval, as the dependent structures are unknown.

Interestingly, our approach for point and interval estimation corresponds to the "exact partial likelihood" for censored data in the presence of ties,

Failure time		Parameters of tables				
Control	6-MP	X	N_1	N_2	M_1	M_2
1, 1		2	21	21	2	40
2, 2		2	19	21	2	38
3		1	17	21	1	37
4, 4		2	16	21	2	35
5, 5		2	14	21	2	33
	6, 6, 6	0	12	21	3	30
	7	0	12	17	1	28
8, 8, 8, 8		4	12	16	4	24
	10	0	8	15	1	22
11, 11		2	8	13	2	19
12, 12		2	6	12	2	16
	13	0	4	12	1	15
15		1	4	11	1	14
	16	0	3	11	1	13
17		1	3	10	1	12
22	22	1	2	7	2	7
23	23	1	1	6	2	5

as suggested in Cox (1972, section 6), a challenging problem in terms of computation. In this connection, we provide a way to perform this exact partial likelihood inference using the Jacobi polynomial. To make a comparison, we also list in Table 7 some other estimators and confidence intervals obtained by the methods suggested by Breslow (1976) and Efron (1977) to approximate the "exact partial likelihood".

Method	Estimator of θ	95% C. I. for θ
Exact Partial Likelihood Inference	5.09	(2.18, 11.91)
Efron Approximation	4.82	(2.15, 10.80)
Breslow Approximation	2.13	(1.42, 3.18)

Table 7 shows that there is a general agreement between our method and Efron's method, and there is also a clear difference between Breslow's method and the others. In particular, Breslow's estimator is not contained in any of the confidence intervals provided by our method and Efron's method; nor do estimators given by our method and Efron method lie

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.

Age (years)		Daily alcohol consumption	
		80+ g	0-79 g
25-34	Case	1	0
	Control	9	106
35-44	Case	4	5
	Control	26	164
45-54	Case	25	21
	Control	29	138
55-64	Case	42	34
	Control	27	139
65-74	Case	19	36
	Control	18	88
75+	Case	5	8
	Control	0	31
	Total sample size: 975		

As indicated in Breslow and Day (1980, p. 146), standard tests show no

evidence of heterogeneity of the odds ratio. Therefore we may assume that $\theta_1 = \cdots = \theta_6$. 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 $0 < \epsilon < \theta$, there exists a constant $M(\epsilon) > 0$ such that

$$P\left(\inf_{\tilde{\theta}:|\tilde{\theta}-\theta|\geq\epsilon} \frac{|\sum_{k=1}^{K} X^{(k)} - \sum_{i=1}^{S}\sum_{k=1}^{K}(1+\tilde{\theta}^{-1}\lambda_i^{(k)})^{-1}|}{\sum_{k=1}^{K} \mathsf{Var}_{\theta}^{(k-1)}(X^{(k)})} > M(\epsilon)\right) \longrightarrow 1, \tag{A.1}$$

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

$$\frac{\sum_{k=1}^{K} X^{(k)} - \sum_{k=1}^{K} \mathsf{E}_{\theta}^{(k-1)}(X^{(k)})}{\sum_{k=1}^{K} \mathsf{Var}_{\theta}^{(k-1)}(X^{(k)})} \xrightarrow{p} 0.$$

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

$$\liminf_{n\to\infty} \inf_{\tilde{\theta}:|\tilde{\theta}-\theta|\geq\epsilon} \frac{|\sum_{k=1}^{K}\sum_{i=1}^{S}(1+\theta^{-1}\lambda_i^{(k)})^{-1} - \sum_{k=1}^{K}\sum_{i=1}^{S}(1+\tilde{\theta}^{-1}\lambda_i)^{-1}|}{\sum_{k=1}^{K} \mathsf{Var}_{\theta}^{(k-1)}(X^{(k)})} \geq \frac{M(\epsilon)}{2}. \tag{A.2}$$

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

$$\sum_{k=1}^{K}\sum_{i=1}^{S}(1+\theta^{-1}\lambda_i^{(k)})^{-1} - \sum_{k=1}^{K}\sum_{i=1}^{S}(1+\tilde{\theta}^{-1}\lambda_i^{(k)})^{-1} = \sum_{k=1}^{K}\sum_{i=1}^{S} \frac{\tilde{\theta}_*^{-2}\lambda_i^{(k)}}{(1+\tilde{\theta}_*^{-1}\lambda_i^{(k)})^2}(\theta - \tilde{\theta})$$

for some $\tilde{\theta}_*$ between θ and $\tilde{\theta}$. Thus we get

$$\text{left side of (A.2)} \geq \liminf_{n\to\infty} \frac{\sum_{k=1}^{K}\sum_{i=1}^{S}(\theta+\epsilon)^{-2}\lambda_i^{(k)}/(1+(\theta-\epsilon)^{-1}\lambda_i^{(k)})^2}{\sum_{k=1}^{K}\sum_{i=1}^{S}\theta^{-1}\lambda_i^{(k)}/(1+\theta^{-1}\lambda_i^{(k)})^2}\epsilon$$
$$\geq \frac{(\theta-\epsilon)(\theta+\epsilon)^{-2}\min(\theta-\epsilon, \frac{1}{\theta-\epsilon})}{\max(\theta, \frac{1}{\theta})}\epsilon,$$

where the second inequality follows from the following elementary inequality

$$\min(a^2/b^2, b^2/a^2)\frac{b^{-2}\lambda_i}{(1+b^{-1}\lambda_i)^2} \leq \frac{a^{-2}\lambda_i}{(1+a^{-1}\lambda_i)^2}$$
$$\leq \max(a^2/b^2, b^2/a^2)\frac{b^{-2}\lambda_i}{(1+b^{-1}\lambda_i)^2},$$

for any $0 < a < \infty$ and any $0 < b < \infty$. Hence (A.2) and (A.1) hold.

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

$$X^{(k)} - \sum_{k=1}^{K}\mathsf{E}_\theta^{(k-1)}(X^{(k)}) = \sum_{k=1}^{K}\sum_{i=1}^{S}\frac{1}{1+\hat{\theta}^{-1}\lambda_i^{(k)}} - \sum_{k=1}^{K}\sum_{i=1}^{S}\frac{1}{1+\theta^{-1}\lambda_i^{(k)}}$$
$$= \sum_{k=1}^{K}\sum_{i=1}^{S}\frac{\theta_*^{-2}\lambda_i^{(k)}}{(1+\theta_*^{-1}\lambda_i^{(k)})^2}(\hat{\theta}-\theta)$$

for some θ_* between θ and $\hat{\theta}$. This, along with the fact

$$\min(\hat{\theta}^2/\theta^2, \theta^2/\hat{\theta}^2)\frac{\theta^{-2}\lambda_i^{(k)}}{(1+\theta^{-1}\lambda_i^{(k)})^2} \leq \frac{\theta_*^{-2}\lambda_i^{(k)}}{(1+\theta_*^{-1}\lambda_i^{(k)})^2}$$
$$\leq \max(\hat{\theta}^2/\theta^2, \theta^2/\hat{\theta}^2)\frac{\theta^{-2}\lambda_i^{(k)}}{(1+\theta^{-1}\lambda_i^{(k)})^2}, \tag{A.3}$$

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

$$\theta^{-1}\left(\sum_{k=1}^{K}\mathsf{Var}_\theta^{(k-1)}(X^{(k)})\right)^{1/2}\cdot(\hat{\theta}-\theta)$$
$$= \frac{\sum_{k=1}^{K}\sum_{i=1}^{S}\theta^{-2}\lambda_i^{(k)}/(1+\theta^{-1}\lambda_i^{(k)})^2}{\sum_{k=1}^{K}\sum_{i=1}^{S}\theta_*^{-2}\lambda_i^{(k)}/(1+\theta_*^{-1}\lambda_i^{(k)})^2}\cdot\frac{X^{(k)}-\sum_{k=1}^{K}\mathsf{E}_\theta^{(k-1)}(X^{(k)})}{\sqrt{\sum_{k=1}^{K}\mathsf{Var}_\theta^{(k-1)}(X^{(k)})}},$$

which converges to $N(0,1)$, via Theorem 3.1.

Finally, we can easily get from (A.3),

$$\min(\hat{\theta}^2/\theta^2, \theta^2/\hat{\theta}^2) \leq \frac{\widehat{\mathsf{Var}}}{\sum_{k=1}^{K}\mathsf{Var}_\theta^{(k-1)}(X^{(k)})} \leq \max(\hat{\theta}^2/\theta^2, \theta^2/\hat{\theta}^2).$$

From this and the fact that $\hat{\theta} \to \theta$ in probability we conclude that (4.4) holds with the normalizing factor replaced by (4.6). □

Proof of Theorem 6.1. To prove the theorem, it suffices to verify the variance stability condition (3.2) in view of Theorem 4.1. Let $D = \{t \in [0,\infty) : F_1(t+) - F_1(t-) = 0\}$. Because of (6.1), $F_2(t)$ is also continuous on the set D. For any $\epsilon > 0$, let $D_\epsilon = \{t \in [0,\infty) : F_1(t+) - F_1(t-) \geq \epsilon\}$ and $D_\epsilon^c = [0,\infty)\backslash(D \cup D_\epsilon)$.

Clearly, for those τ_k in D, $M_1^{(k)} = 1$, as there is no tied observation. When $M_1^{(k)} = 1$, the noncentral hypergeometric distribution becomes Bernoulli distribution with success probability $\theta N_1^{(k)}/(\theta N_1^{(k)} + N_2^{(k)})$. Therefore, for the conditional variance of $X^{(k)}$, we get

$$\frac{1}{n}\sum_{k:\,\tau_k\in D}\mathsf{Var}_\theta^{(k-1)}(X^{(k)}) = \frac{1}{n}\sum_{k:\,\tau_k\in D}\frac{\theta N_1^{(k)}N_2^{(k)}}{\left(\theta N_1^{(k)} + N_2^{(k)}\right)^2}$$
$$= \frac{1}{n}\int_D \frac{\theta N_1(t)N_2(t)}{(\theta N_1(t) + N_2(t))^2}\mathrm{d}J(t),$$

which converges to a nonrandom constant by the law of large numbers, where $J(t) = \#\{i : \tilde{T}_i \leq t, \Delta_i = 1\}$.

We next deal with set D_ϵ. By definition, the number of points in D_ϵ, denoted by $r = r(\epsilon)$, is finite. Denote all the points in D_ϵ to be t_1^ϵ, t_2^ϵ, ... , t_r^ϵ. Then, as $n \to \infty$, the conditional variances of all tables constructed on the times t_1^ϵ, ... , t_r^ϵ will all go to ∞. Therefore, from Kou and Ying (1996), the conditional variances can be approximated by

$$\frac{1}{n}\sum_{k:\,\tau_k\in D_\epsilon}\mathsf{Var}_\theta^{(k-1)}(X^{(k)})$$
$$= \frac{1}{n}\sum_{j=1}^{r}\left(\frac{1}{X_\epsilon^{(j)}} + \frac{1}{M_1\left(t_j^\epsilon\right) - X_\epsilon^{(j)}} + \frac{1}{N_1\left(t_j^\epsilon\right) - X_\epsilon^{(j)}}\right.$$
$$\left. + \frac{1}{N_2\left(t_j^\epsilon\right) - M_1\left(t_j^\epsilon\right) - N_1\left(t_j^\epsilon\right) + X_\epsilon^{(j)}}\right)^{-1} + o_p(1),$$

where $X_\epsilon^{(j)} = \#\{i \leq n_1 : \tilde{T} = t_j^\epsilon \text{ and } \Delta_i = 1\}$. This, again by the law of large numbers, converges to a nonrandom constant.

Finally, from (A.3) we can bound the conditional variance of $X^{(k)}$ at θ

with that at $\theta = 1$ to get

$$\begin{aligned}\frac{1}{n}\sum_{k:\,\tau_k\in D_\epsilon^c}\mathrm{Var}_\theta^{(k-1)}(X^{(k)}) &\le \max(\theta,\,\frac{1}{\theta})\frac{1}{n}\sum_{k:\,\tau_k\in D_\epsilon^c}\frac{N_1^{(k)}N_2^{(k)}M_1^{(k)}M_2^{(k)}}{\left(N^{(k)}\right)^2\left(N^{(k)}-1\right)}\\ &\le \max(\theta,\,\frac{1}{\theta})\frac{1}{n}\sum_{k:\,\tau_k\in D_\epsilon^c}M_1^{(k)}\\ &\le \max(\theta,\,\frac{1}{\theta})\int_{D_\epsilon^c}(\mathrm{d}F_1(t)+\mathrm{d}F_2(t))+o_p(1),\end{aligned}$$

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 $n^{-1}\sum_k \mathrm{Var}_\theta^{(k-1)}(X^{(k)})$ 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

1. Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.
2. Breslow (1976), Regression analysis of the log odds ratio: a method for retrospective studies, *Biometrics*, **32**, 409-416.
3. Breslow (1981), Odds ratio estimators when the data are sparse, *Biometrika*, **68**, 73-84.
4. Breslow (1996), Statistics in epidemiology: the case-control studies, *J. Amer. Statist. Assoc.*, **91**, 14-28.
5. Breslow, N. E. and Cologne, J. (1986), Methods of estimation in log odds ratio regression models, *Biometrics*, **42**, 949-954.
6. Breslow, N. E. and Day, N. E. (1980), *Statistical Methods in Cancer Research, Vol. I, The Design and Analysis of Case-Control Studies*, Lyon: IARC.
7. Breslow, N. E. and Day, N. E. (1987), *Statistical Methods in Cancer Research, Vol. II, The Design and Analysis of Cohort Studies*, Lyon: IARC.
8. Chow, Y. S., and Teicher, H. (1988), *Probability Theory*, New York: Springer-Verlag.
9. Cox, D.R. (1972), Regression models and life tables (with discussion), *J. R. Statist. Soc. B*, **74**, 187-220.
10. Cox, D.R. (1975), Partial likelihood, *Biometrika*, **62**, 269-276.
11. Cox, D.R. and Snell, E.J. (1989), *Analysis of Binary Data*, London: Chapman and Hall.
12. Cytel Inc. (2003), *StatXact: A Statistical Package for Exact Nonparametric inference, Version 6*, Cambridge, MA.

13. de la Peña, V.H. and Giné, E. (1999), *Decoupling: From Dependence to Independence*, New York: Springer-Verlag.
14. Efron (1977), The efficiency of Cox's likelihood function for censored data, *J. of Amer. Stat. Assoc.*, **72**, 557-65.
15. Freireich, E. J. *et al.* (1963), The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia. *Blood*, **21**, 699-0716.
16. Fleming, T. R. and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.
17. Gart, J. (1970), Point and interval estimation of the common odds ration in the combination of 2×2 tables, *Biometrika*, **57**, 471-475.
18. Gehan, E. A. (1965), A generalized Wilcoxon test for comparing arbitrarily single-censored samples, *Biometrika*, **52**, 203-23.
19. Hauck, W. W. (1989), Odds ratio inference from stratified samples, *Comm. in Statis.–Theory and Method*, **18**, 767-800.
20. Johnson, L. D. and Riess, R. D. (1982), *Numerical Analysis*, New York: Addison-Wesley.
21. Kou, S. G. and Ying, Z. (1996), Asymptotics for a 2×2 table with fixed margins, *Statistica Sinica*, **6**, 809-829.
22. Kwapień, S. and Woyczyński (1992), *Random Series and Stochastic Integrals: Single and Multiple*, Boston: Birkhäuser.
23. Liang, K. Y. and Self, S. (1985), Tests for homogeneity of odds ratio when the data are sparse. *Biometrika*, **72**, 353-358.
24. Loève, M. (1977), *Probability Theory Vol. I*, New York: Springer-Verlag.
25. Mantel, N. (1963), Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J. of Amer. Stat. Assoc.* **58**, 690-700.
26. Mantel, N. and Haenszel, W. (1959), Statistical aspects of the analysis of data from retrospective studies of disease, *J. Nat. Cancer Inst.* **22**, 719–748.
27. MathSoft (1995), *S-plus Manuals*, MathSoft: Seattle
28. McCullagh, P. (1984), On the elimination of nuisance parameters in the proportional odds model, *J. of the Royal Stat. Society, Series B*, **46**, 250-256.
29. McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman and Hall.
30. Mehta, C. R., Patel, N. R., and Gray, R. (1985), Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables, *J. of Amer. Stat. Assoc.*, **80**, 969-973.
31. Miller, R. G. (1981), *Survival Analysis*, New York: Wiley.
32. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika*, **64**, 175-188.
33. Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer-Verlag.
34. Robins, J., Breslow, N. and Greenland, S. (1986), A Mantel-Haenszel variance consistent under both large strata and sparse data limiting models, *Biometrics*, **42**, 311-323.
35. Santner, T. and Duffy. D. (1989), *The statistical Analysis of Discrete Data*, New York: Springer-Verlag.
36. Strawderman, R. and Wells, M. (1998). Approximately exact inference for

the common odds ratio in several 2 X 2 tables, with discussion. *J. of Amer. Stat. Assoc.*, **93**, 1294 - 1320.
37. Szegö, G. (1959), *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Pub., Vol. **23**, Providence: Amer. Math. Soc.
38. Tarone, R. E. (1985), On heterogeneity tests based on efficient scores, *Biometrika*, **72**, 91-95.
39. Tarone, R. E. and Ware, J. (1977), On distribution-free tests for equality of survival distributions, *Biometrika*, **64**, 156-160.
40. Wolfram, S. (1991), *Mathematica: A System for Doing Mathematics by Computer*, New York: Andison-Wesley.
41. Woolf, B. (1955), On estimating the relationship between blood group and disease, *Ann. Human Genet.*, **19**, 251-253.
42. Zelen, M. (1971), The analysis of several 2×2 contingency tables. *Biometrika*, **58**, 129-37.

A NEW TEST OF SYMMETRY ABOUT AN UNKNOWN MEDIAN

WEIWEN MIAO

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

YULIA R. GEL

Department of Statistics and Actuarial Science, University of Waterloo Waterloo, ON, Canada N2L 3G1

JOSEPH L.GASTWIRTH[3]

Department of Statistics, George Washington University Washington, DC, 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-Solomons measure of asymmetry by corresponding robust estimators; the median and mean 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 Cabilio and Masaro (1996) and Mira (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 lambda 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 μ,