Research Memorandum No.385 July 30, 1990
The Institute of Statistical Mathematics

A Simple Recursive Algorithm for the Exact Confidence Limits for the Common Odds Ratio in a Series of 2×2 Tables

Hirofumi Takagi
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Summary

A recursive algorithm is provided for computing the exact confidence limits for the common odds ratio in several 2×2 contingency tables in order to elucidate a dichotomous risk factor without the influences of the confounding variables. The computing time by our algorithm increases as the order of sample size square, which is equivalent to that of the network algorithm provided by Mehta et al (1985, *JASA*). Our algorithm, however, is simpler and more programmatic rather than the network algorithm. A simple recursive algorithm can easily and rapidly give the exact confidence limits for the common odds ratio for all situations between sparse data settings and large-stratum settings where asymptotic approaches may be available but suspect.

Key Words: Recursive algorithm; Mantel-Haenszel method; Exact confidence limits; Common odds ratio; Case-control study; Epidemiology;

1. Introduction

In the analyses of the epidemiologic studies, especially in case-control studies, the data are often stratified into a series of 2×2 tables to elucidate a dichotomous risk factor without the influences of the confounding variables. Therefore, we usually try to compute the estimate and confidence limits for ψ, the common odds ratio in K 2×2 independent contingency tables. Several asymptotic methods have been proposed to this problem.

The Mantel-Haenszel (1959) estimator, $\hat{\psi}_{MH}$ is ordinarily used to estimate ψ. To construct the confidence limits for ψ, the asymptotic variance of the Mantel-Haenszel estimator was given by Hauck (1979) based on "large-stratum" sets wherein the number of tables remained fixed but individual cell sizes increased without bound. Breslow (1981) proposed one variance for "sparse-data" settings where the cell sizes remained bounded but the number of tables increased. Robins et al (1986) provided a good asymptotic variance for both sparse-data and large-stratum settings. We construct an approximate confidence interval by the "ln-method" that is constructed by using $\log \hat{\psi}_{MH}$ and its asymptotic variance, and the approximate $1 - \alpha$ confidence limits are given by

$$\hat{\psi}_{MH} \exp\left[\pm \chi_\alpha \cdot \sqrt{\widehat{\mathrm{var}}(\log \hat{\psi}_{MH})}\right],$$

where χ^2_α is the upper α point of the chi-square distribution with one degree of freedom.

Another approximate methods are derived from the test statistics such as the extended Cornfield method by Gart (1970, 1971) that performs well for large-stratum sets but poorly for sparse-data sets, while Sato (1990) proposed a good method for both sparse-data and large-stratum settings.

However, we can also construct an exact confidence interval based on the conditional distribution (Zelen, 1971), which guarantees the desired coverage probabilty. Though the computational complexity exists in estimating the exact confidence limits, Mehta et al (1985) provided the network algorithm to overcome the time consumption even for sparse-data settings, of which the computational time increases as the order of $(nK)^2$. The network algorithm, however, is not essential to compute the exact confidence limits although the idea may be relevant. In fact, some recursive process is essential to obtain the exact confidence interval.

In this paper, we provide a simple recursive algorithm to compute an exact confidence

interval for ψ, of which computational time increases as the order of $(nK)^2$ that is equivalent to that of the network algorithm, while our algorithm is simpler and easier to complete a computer programme.

2. Exact Inference for the Common Odds Ratio

Consider K 2×2 tables with pairs of independent binomial observations $(x_k,\ y_k)$ with denominators $(n_k,\ m_k)$ and success probabilities $(p_{1k},\ p_{2k})$ for $k = 1, \ldots, K$. In case-control studies x_k and y_k persons are exposed in n_k cases and m_k controls respectively in the kth stratum.

To measure the disease-exposure association the odds ratio $\psi_k = (p_{1k}q_{2k})/(p_{2k}q_{1k})$ is available, where $q_{1k} = 1 - p_{1k}$ and $q_{2k} = 1 - p_{2k}$ in the kth table. Usually assumed $\psi_k = \psi$ for all k, where ψ is the common odds ratio.

The "exact" estimator and confidence limits for ψ are derived from the conditional distribution(Gart, 1970, and Zelen, 1971). Given the table marginals n_k, m_k and $t_k = x_k + y_k$, the conditional distribution is the noncentral hypergeometric distribution for x_k,

$$f(x_k|t_k, \psi) = \frac{W(k, x_k)\psi^{x_k}}{\sum_{r_k} W(k, r_k)\psi^{r_k}}, \tag{1}$$

where

$$W(k, r_k) = \binom{n_k}{r_k}\binom{m_k}{t_k - r_k},$$

for

$$l_k = \max(0, t_k - m_k) \leq r_k \leq \min(n_k, t_k) = u_k.$$

The conditional maximum likelihood estimator $\hat{\psi}_c$ can be obtained by the following equation with an iterative method.

$$\sum_{k=1}^{K} x_k = \sum_{k=1}^{K} \sum_{r_k=l_k}^{u_k} r_k\ f(r_k|t_k, \hat{\psi}_c). \tag{2}$$

Because of time consuming against the truth, most of epidemiologists and biostatistian prefer the Mantel-Haenszel estimator $\hat{\psi}_{MH}$ to CMLE $\hat{\psi}_c$. However, the CMLE can be easily computed without time spending by 16 bit micro-computer (Satten and Kupper, 1990, and Takagi, 1990).

Let $L = \sum_k l_k$ and $U = \sum_k u_k$, the conditional distribution of $s = \sum_k x_k$ is given (Zelen, 1971) by

$$g(s|\psi) = C_s \psi^s \Big/ \sum_{r=L}^{U} C_r \psi^r, \tag{3}$$

where

$$C_s = \sum_{R(s)} \prod_{k=1}^{K} W(k, r_k),$$

and $R(s)$ is the set of all K-fold partitions of all $\sum_k r_k = s$.

Exact $100(1-\alpha)\%$ confidence limits for ψ are given by $\{\hat{\psi}_L,\ \hat{\psi}_U\}$ such that

$$\sum_{z=s}^{U} g(z|\hat{\psi}_L) = \frac{\alpha}{2}, \tag{4}$$

and

$$\sum_{z=L}^{s} g(z|\hat{\psi}_U) = \frac{\alpha}{2}. \tag{5}$$

Note that $\hat{\psi}_L = 0$ if $s = L$, and $\hat{\psi}_U = \infty$ if $s = U$.

To estimate exact confidence limits, we need an efficient way to compute coefficients C_s and also must take a measure against numerical overflow. If these problems are resolved, we can easily compute exact confidence limits using an iterative method such as a Newton-Raphson method or a binary search.

3. A Simple Recursive Algorithm

First of all, we should compute $W(k, r_k)$ for all k and r_k. In general, however, we will use $\log W(k, r_k)$ that can be easily obtained instead of $W(k, r_k)$ as explained later.

A simple recursive algorithm for the computation of coefficients C_s is as follows;

[1] Let $C(1, r_1) = W(1, r_1)$ for $l_1 \le r_1 \le u_1$.

[2] Let $L_k = \sum_{j=1}^{k} l_j$ and $U_k = \sum_{j=1}^{k} u_j$. Then compute the following equation.

$$C(k+1, z_{k+1}) = \sum_{z_k + r_{k+1} = z_{k+1}} C(k, z_k) \cdot W(k+1, r_{k+1}), \tag{6}$$

where $L_k \le z_k \le U_k$, and $l_{k+1} \le r_{k+1} \le u_{k+1}$.

[3] Repeat [2] for $k = 2, \ldots, K$.

When all computations are finished, then we can obtain $C_z = C(K, z_K)$.

Hence, the maximum computation amount described above is given by

$$\sum_{k=2}^{K}\left[\sum_{j=1}^{k-1} n_j + 1\right](n_k+1) \leq (K-1)(n+1) + \frac{(K-1)K}{2}n(n+1), \tag{7}$$

that follows the order of $(nK)^2$, where $n = \max(n_1, \ldots, n_K)$.

In practical applications, we may encounter to a technical difficulty while computing $W(k, r_k)$, $C(k, z_k)$ and ψ^s if the above procedure will be operated originally, that is the problem about numerical overflow. In order to avoid this problem, we must handle the natural logarithm of these values such that let $lW(k, r_k) = \log W(k, r_k)$ and $lC(k, z_k) = \log C(k, z_k)$. In fact, $lW(k, r_k)$ can be easily computed by

$$\begin{aligned} lW(k, r_k) &= \log\Gamma(n_k+1) + \log\Gamma(m_k+1) - \log\Gamma(r_k+1) - \log\Gamma(n_k - r_k + 1) \\ &\quad - \log\Gamma(t_k - r_k + 1) - \log\Gamma(m_k - t_k + r_k + 1), \end{aligned}$$

where $\log\Gamma(\cdot)$ is the log-gamma function.

If a mathematical function library for FORTRAN contains the log-gamma function, we can use it, but if not, we should compute it. Fortunately, however, since we use only integer numbers, it is not difficult to compute the values of the function by using the following equations; $\log\Gamma(n+1) = \log n + \log\Gamma(n)$ and $\log\Gamma(1) = 0$. In practice, we had better obtain the values of $\log\Gamma(n)$ in advance for n is between 1 and large enough for the computations.

To avoid the overflow problem, we may use the following equation in stead of (6);

$$\begin{aligned} lC(k+1, z_{k+1}) &= \log\left[\sum_{z_k + r_{k+1} = z_{k+1}} \exp\{\, lC(k, z_k) + lW(k+1, r_{k+1}) - \Delta(z_{k+1})\}\right] \\ &\quad + \Delta(z_{k+1}), \end{aligned} \tag{8}$$

where $\Delta(z_{k+1}) = \max[\, lC(k, z_k) + lW(k+1, r_{k+1});\ z_{k+1} = z_k + r_{k+1},\ L_k \leq z_k \leq U_k,\ l_{k+1} \leq r_{k+1} \leq u_{k+1}]$.

Since $\Delta(z_{k+1})$ is evaluated and revised in each step in practical computations, it is not necessary to detect the maximum value for all $lC(k, z_k) + lW(k+1, r_{k+1})$ at the beginning of the procedure.

When computing (3), (4), and (5), $C_r\psi^r$ may cause the numerical overflow. Therefore, we take a measure against the problem. In order to avoid it, rearrange the left term in (4) for $\hat{\psi}_L$ as follows;

$$\frac{\sum_{z=s}^{U} \exp[\log C_z + z\log\hat{\psi}_L - \Delta]}{\sum_{r=L}^{U} \exp[\log C_r + r\log\hat{\psi}_L - \Delta]} = \frac{\alpha}{2}. \tag{9}$$

where $\Delta = \max[\log C_r + r\log\hat{\psi}_L;\ L \le r \le U\ ]$, and the same manner can be adopted to the estimation for $\hat{\psi}_U$.

Clearly, we had better obtain $\log C_r$ instead of coefficients C_r for computing (9) for large K and/or large samples.

4. Numerical Example

In this section, we show a practical procedure for the algorithm described above by using a simple example. Suppose $K = 3$, $n_k = m_k = t_k = 2$ for $k = 1, 2$, and 3, and $s = 3$. Consider three tables as follows;

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}.$$

Then we can easily compute $W(k,0) = 1$, $W(k,1) = 4$, and $W(k,2) = 1$ for all k.

Table 1. The first computation procedure for coefficients C_z.

		r_2	0	1	2	$(z_1 + r_2)$	
z_1	$C(1,z_1)$	$W(2,r_2)$	1	4	1	z_2	$C(2,z_2)$
0	1		1×1	—	—	0	01
1	4		4×1	1×4	—	1	08
2	1		1×1	4×4	1×1	2	18
–	–		—	1×4	4×1	3	08
–	–		—	—	1×1	4	01

Table 1 shows the process of computing $C(2,z_2)$. In this case, we only consider two 2×2 tables. First of all, let $C(1,z_1) = W(1,r_1)$ for $z_1 = r_1 = 0, 1$, and 2. z_1 and $C(1,z_1)$ are written in the first and second columns in the table, and r_2 and $W(2,r_2)$ are in the top and second rows. Calculate $C(1,z_1) \times W(2,r_2)$ for all z_1 and r_2, and put them in the proper locations as shown in Table 1. Note that the multiplications in each column move downward step by step according to $z_1 + r_2$, say z_2.

Finally, we sum upp the values in each row, respectively. Then we can obtain $C(2, z_2)$ for $z_2 = 0, 1, 2, 3$, and 4.

Table 2. The second computation procedure for coefficients C_z.

		r_3	0	1	2	$(z_2 + r_3)$	
z_2	C(2,z_2)	$W(3, r_3)$	1	4	1	z_3	$C(3, z_3)$
0	01		1×1	—	—	0	01
1	08		8×1	1×4	—	1	12
2	18		18×1	8×4	1×1	2	51
3	08		8×1	18×4	8×1	3	88
4	01		1×1	8×4	18×1	4	51
–	–		—	1×4	8×1	5	12
–	–		—	—	1×1	6	01

For three 2 × 2 tables, we can use the results of two 2 × 2 tables. Table 2 shows the process of computing $C(3, z_3)$ for $z_3 (= z_2 + r_3)$. At glance, clearly, the computing procedures are the same as those shown in Table 1. Then we can easily obtain $C(3, z_3)$ for $z_3 = 0, 1, 2, 3, 4, 5$, and 6.

If $K \geq 4$, all we need is to repeat the same process recursively until desired results are obtained.

Thus, if $s = 3$ and $\psi = 1$, we can easily compute the probability as follows;

$$g(3|\psi = 1) = \frac{C_3}{\sum_{z=0}^{6} C_z} = \frac{88}{1 + 12 + 51 + 88 + 51 + 12 + 1} = 0.4074$$

Exact confidence limits for ψ can be obtained by the same way using (4) and (5), though an available computer programme should be necessary.

5. Discussion

We have provided a simple recursive algorithm for estimating exact confidence limits for the common odds ratio in several 2 × 2 tables. Our algorithm demands time consumption as the order of $(nK)^2$ that is equivalent to that of the network algorithm by Mehta et al (1985). However, our algorithm is quite simpler and more easily programmatic rather than the network algorithm.

In practice, though approximate methods for confidence interval may be available in many cases for large-stratum settings and/or even for sparse-data settings, as Mehta et al (1985) discussed in details, one would prefer to obtain exact confidence limits if asymptotic approximations are questioned by the skewness of table marginals and/or large ψ.

Moreover, since approximate methods are asymptotic, if we can obtain exact confidence limits, then we will have additional exact information about the common odds ratio, and therefore evaluate efficiently the risk of dichotomous variable of interest. Even if exact methods consume time more than approximate methods, what is important is to be able to obtain exact confidence limits for ψ with the desired coverage probability.

The FORTRAN programme for our algorithm is available from the author on request.

Acknowledgement

The author wishes to thank Professor Takemi Yanagimoto, the Institute of Statistical Mathematics, for his relevant suggestions and helpful comments.

Refferences

Breslow, N.E. (1981). Odds ratio estimators when the data are sparse. *Biometrika* **68**, 73–84.

Gart, J.J. (1970). Point and interval estimation of the common odds ratio in the combination of 2 × 2 tables with fixed marginals. *Biometrika* **57**, 471–475.

Gart, J.J. (1971). The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification. *Review of the International Statistical Institute* **39**, 148–169.

Hauck, W.W. (1979). The large sample variance of the Mantel-Haenszel estimator of a common odds ratio. *Biometrics* **35**, 817–819.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719–748.

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. *Journal of the American Statistical Association* **80**, 969–973.

Robins, J.M., Breslow,N.E., and Greenland, S. (1986). Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting model. *Biometrics* **42**, 311–323.

Sato, T. (1990). Confidence limits for the common odds ratio based on the asymptotic distribution of the Mantel-Haenszel estimator. *Biometrics* **46**, 71–80.

Satten, G. A. and Kupper L.L. (1990). Continued fraction representation for expected cell counts of a 2×2 table; A rapid and exact method for conditional maximum likelihood estimation. *Biomterics* **46**, 217–223.

Takagi, H. (1990). Programme on the conditional maximum likelihood estimator for the common odds ratio and the AIC for the analysis of K 2×2 tables. *Computer Science Monographs*, submitted.

Zelen, M. (1971). The analysis of several 2×2 contingency tables. *Biometrika* **58**, 129–137.

DOCUMENTATION PAGE

Research Memorandum Number	No.385	Performing Organization Name Division of Research Information The Institute of Statistical Mathematics
Report Date	July 30, 1990	

Title (and Subtitle)

A Simple Recursive Algorithm for the Exact Confidence Limits for the Common Odds Ratio in a Series of 2×2 Tables

Author(s) and Address

Hirofumi Takagi
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Number of Pages	AMS 1980 Subject Classifications	Distribution Statement (of this Res. Memo.)
10		

Key Words

Recursive algorithm; Mantel-Haenszel method; Exact confidence limits; Common odds ratio; Case-control study; Epidemiology;

Abstracts (Continue on Reverse Side if Necessary)

A recursive algorithm is provided for computing the exact confidence limits for the common odds ratio in several 2 × 2 contingency tables in order to elucidate a dichotomous risk factor without the influences of the confounding variables. The computing time by our algorithm increases as the order of sample size square, which is equivalent to that of the network algorithm provided by Mehta et al(1985, *JASA*). Our algorithm, however, is simpler and more programmatic rather than the network algorithm. A simple recursive algorithm can easily and rapidly give the exact confidence limits for the common odds ratio for all situations between sparse data settings and large-stratum settings where asymptotic approaches may be available but suspect.