Statistics and Probability Letters 82 (2012) 1990–1993

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

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

On closeness of the Mantel–Haenszel estimator and the profile likelihood based estimator of the common odds ratio from multiple 2×2 tables

Buddhananda Banerjee, Atanu Biswas *

Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700 108, India

A R T I C L E   I N F O

Article history:
Received 11 January 2012
Received in revised form 13 June 2012
Accepted 16 June 2012
Available online 23 June 2012

Keywords:
Odds ratio
Mantel–Haenszel estimator
Maximization
Profile likelihood

A B S T R A C T

In this paper, we show the difference in the variance of the Mantel–Haenszel estimator and the profile maximum likelihood estimator of order $O((kn)^{-3})$ or less, where the individual variances are of order $O(n^{-1})$. Here n is the minimum number of samples available for any treatment in any group. Numerical results are also reported.

1. Introduction

The odds ratio for 2×2 tables is extensively used in biomedical studies including clinical trials. Consider the following typical 2×2 table.

		Treatment		Total
		A	B	
Response	Success	X	Y	$X + Y$
	Failure	$n_A - X$	$n_B - Y$	$n_A + n_B - X - Y$
	Total	n_A	n_B	$N = n_A + n_B$

Suppose, marginally $X \sim Bin(n_A, p_A)$ and $Y \sim Bin(n_B, p_B)$. Then the odds ratio is denoted by

$$\psi = \frac{p_A(1-p_B)}{p_B(1-p_A)}.$$

Usually ψ is estimated by $\widehat{\psi} = \frac{\widehat{p}_A(1-\widehat{p}_B)}{\widehat{p}_B(1-\widehat{p}_A)} = \frac{X(n_B-Y)}{Y(n_A-X)}$, with $\widehat{p}_A = X/n_A$ and $\widehat{p}_B = Y/n_B$. The distribution of $\log \widehat{\psi}$ is often approximated by a normal distribution with mean $\log \psi$ and variance σ^2, which is estimated by

$$\widehat{\sigma}^2 = \frac{1}{X} + \frac{1}{n_A - X} + \frac{1}{Y} + \frac{1}{n_B - Y}. \tag{1.1}$$

* Corresponding author. Tel.: +91 33 25772088; fax: +91 33 25773104.
E-mail addresses: buddhananda_r@isical.ac.in (B. Banerjee), atanu@isical.ac.in (A. Biswas).

0167-7152/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2012.06.013

See Breslow (1981). Anscombe (1956) suggested to add 0.5 to all the 4 cells to incorporate the cases $X = 0$ or $Y = 0$ or $n_A - X = 0$ or $n_B - Y = 0$.

For a series of k independent 2×2 tables with the common odds ratio ψ, the odds ratio is usually estimated by the Mantel–Haenszel (MH) estimator (Mantel and Haenszel, 1959), or by the profile maximum likelihood estimator (PMLE). These are discussed in Section 2. Variance and the estimate of variance of the MH estimator is well studied in the literature (see Breslow and Liang (1982), Falanders (1985), Robins et al. (1986), Phillps and Holland (1987) and Pigeot (1991)). Pigeot (1990) had also studied a class of asymptotically efficient non-iterative estimators of a common odds ratio. Variance of PMLE is available from likelihood by inverting the Fisher information matrix (see Gart (1992) and Murphy and van der Vaart (2000)). Although it is proven that asymptotic efficiency of PMLE is more than that of the MH estimator (Tarone et al., 1983; Hauck, 1979), we are interested to know whether this gain is significant over the computational intensity of PMLE. In Section 3, we obtain the asymptotic closeness of variances of the MH estimator and PMLE. We also present a detailed simulation study for small sample behavior of their closeness. Section 4 concludes.

2. Multiple tables: the MH estimator and PMLE

Suppose there are k 2×2 tables. If the ith table has the cell frequencies X_i, $n_{Ai} - X_i$, Y_i and $n_{Bi} - Y_i$, $i = 1, \ldots, k$, the MH-estimator is given by

$$\widehat{\psi}_{\text{MH}} = \frac{\sum_{i=1}^{k} \left(\frac{n_{Ai}n_{Bi}}{N_i}\right) \widehat{q}_{Ai}\widehat{p}_{Bi}\widehat{\psi}_i}{\sum_{i=1}^{k} \left(\frac{n_{Ai}n_{Bi}}{N_i}\right) \widehat{q}_{Ai}\widehat{p}_{Bi}} = \frac{\sum_{i=1}^{k} X_i(n_{Bi} - Y_i)/N_i}{\sum_{i=1}^{k} Y_i(n_{Ai} - X_i)/N_i} = \frac{\sum_{i=1}^{k} R_i}{\sum_{i=1}^{k} S_i}, \tag{2.1}$$

with $N_i = n_{Ai} + n_{Bi}$ and $R_i = X_i(n_{Bi} - Y_i)/N_i$, $S_i = Y_i(n_{Ai} - X_i)/N_i$, and $\widehat{\psi}_i = \widehat{p}_{Ai}\widehat{q}_{Bi}/(\widehat{q}_{Ai}\widehat{p}_{Bi})$ is the estimate of the odds ratio ψ for the ith table. The variance of this estimator, assuming independent binomials, has been obtained by different authors. Note that $\widehat{\psi}_{\text{MH}}$ is asymptotically normal with mean ψ and variance $\sigma^2_{\text{MH}} = \left\{\lim_{k\to\infty} \sum_{i=1}^{k} \text{Var}((R_i - \psi S_i)/k)\right\} / \left\{\lim_{k\to\infty} \sum_{i=1}^{k} E(S_i)/k\right\}^2$, which is consistently estimated by $\widehat{\sigma}^2_{\text{MH}} = \sum_{i=1}^{k} (R_i - \widehat{\psi} S_i)^2 / \left(\sum_{i=1}^{k} S_i\right)^2$. The closed form expression of the asymptotic variance of the MH estimator can be obtained as

$$\text{Var}(\widehat{\psi}_{\text{MH}}) = \psi^2 \frac{\sum_{i=1}^{k} \mu_i^2/\omega_i}{\left(\sum_{i=1}^{k} \mu_i\right)^2} \tag{2.2}$$

where $\mu_i = \left(\left(\frac{n_{Ai}n_{Bi}}{N_i}\right) q_{Ai}p_{Bi}\right)$ and $\psi^2/\omega_i = \text{Var}(\widehat{\psi}_i)$ (see Biswas and Hwang (2011)), where $1/\omega_i$ can be estimated using (1.1).

As an alternative, maximization of profile likelihood can also be done for estimating the common odds ratio across the k tables (see Bohning et al. (2008)). Consider the likelihood

$$L(\mathbf{p}_A, \mathbf{p}_B|\mathbf{X}, \mathbf{Y}) = \prod_{i=1}^{k} \binom{n_{Ai}}{X_i} p_{Ai}^{X_i}(1 - p_{Ai})^{n_{Ai}-X_i} \binom{n_{Bi}}{Y_i} p_{Bi}^{Y_i}(1 - p_{Bi})^{n_{Bi}-Y_i},$$

where $\mathbf{p}_A = (p_{A1}, p_{A2}, \ldots, p_{Ak})$, $\mathbf{p}_B = (p_{B1}, p_{B2}, \ldots, p_{Bk})$ and $\mathbf{X} = (X_1, X_2, \ldots, X_k)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$. Now denote $\theta_{Ai} = p_{Ai}/(1 - p_{Ai})$, $\theta_{Bi} = p_{Bi}/(1 - p_{Bi})$ and assume that $\psi = \theta_{Ai}/\theta_{Bi}$, for $i = 1, \ldots, k$. Replacing $\theta_{Ai} = \psi\theta_{Bi}$ in the log-likelihood and taking the partial derivative with respect to θ_{Bi} and equating to zero we get the feasible solution of θ_{Bi} as a function of ψ. Thus, the profile log-likelihood is a function of the unknown parameter ψ only. Differentiating the profile log-likelihood with respect to ψ and equating it to zero we get an iterative solution for ψ as $\psi = \phi(\psi)$. The iteration becomes faster by adding some specified common term to both sides of the iterative equation. See Bohning et al. (2008, p. 123) for details.

In fact, it is well-known that for the iterative solution of the PMLE, $\widehat{\psi}_{\text{PMLE}}$, the first iteration gives the MH estimator if the initial value of ψ is taken to be 1. See Bohning et al. (2008, pp. 124–125) for details. Note that $\widehat{\psi}_{\text{PMLE}}$ uses $\widehat{p}_{Bi}(\psi)$'s which are the PMLEs of p_B for the ith group assuming common odds ratio ψ. Murphy and van der Vaart (2000) has studied consistency and asymptotic normality of PMLE in general and the asymptotic variance of the profile likelihood based estimator is

$$\text{Var}(\widehat{\psi}_{\text{PMLE}}) = \psi^2 / \sum_{i=1}^{k} \omega_i. \tag{2.3}$$

See Hauck (1979, p. 818) and Gart (1992).

Table 1
Difference between asymptotic variance of $\widehat{\psi}_{\text{MH}}$ and $\widehat{\psi}_{\text{PMLE}}$ [order 10^{-8}].

	k	3	5	7	10	12	15	20	30
n	30	11.71	4.31	0.71	0.96	1.17	1.06	0.58	0.33
	50	3.71	1.62	0.53	1.08	0.39	0.30	0.32	0.18
	80	2.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
	100	2.25	0.73	0.37	0.23	0.23	0.16	0.12	0.07
	120	1.82	0.54	0.26	0.20	0.17	0.12	0.10	0.05
	150	1.26	0.47	0.27	0.14	0.12	0.08	0.08	0.05

3. Closeness of the MH estimator and PMLE for the odds ratio

We have already discussed that both the MH estimator and PMLE are used in intention to estimate the common odds ratio. It is interesting to study how the results from these two commonly used approaches for estimating the common odds ratio vary. In fact that might give us a practical guideline for using one particular method of estimation in some specific situation. It is proven by Tarone et al. (1983) that PMLE is asymptotically more efficient than the non-iterative MH estimator. Still we are interested to know how much efficiency is lost by the use of the MH estimator instead of using PMLE, because the MH estimator is of closed form and simple to compute, whereas PMLE is iterative and so more time consuming in comparison. Interestingly, we have the following theorem whose proof is given in the Appendix.

Theorem 1. *Under equality of the odds ratio across k tables, Var*$(\widehat{\psi}_{\text{MH}}) - $*Var*$(\widehat{\psi}_{\text{PMLE}})$ *is of order less than or equal to* $O((kn)^{-3})$, *where* $n = \min\{n_{Ai}, n_{Bi}, i = 1, \dots, k\}$.

Note that $\text{Var}(\widehat{\psi}_{\text{MH}})$ and $\text{Var}(\widehat{\psi}_{\text{PMLE}})$ are of order $O(n^{-1})$. To supplement the result obtained in Theorem 1, we carry out detailed simulation studies to understand how large sample sizes are required to get practically equivalent results. Table 1 gives numerical results for the common odds ratio $\psi = 1.2$, for varying number of groups from ($k =$) 3 to 30 and for different sample sizes from 30 to 150. For each specified sample size (n) and for each value of k we considered k equidistant points in [0.20, 0.75] as the values of p_{Bi}s for $i = 1, \dots, k$. We then obtain the corresponding $\widehat{p}_{Ai}$s such that the odds ratio $\psi = 1.2$ is kept fixed. For each such specification we calculate $\text{Var}(\widehat{\psi}_{\text{MH}})$ from (2.2) and $\text{Var}(\widehat{\psi}_{\text{PMLE}})$ from (2.3). We observe that although the asymptotic variance of $\widehat{\psi}_{\text{PMLE}}$ is less than that of $\widehat{\psi}_{\text{MH}}$ they are remarkably close for moderate sample sizes. The differences being very small, the MH estimator can be preferred over the PMLE to avoid computational hazards.

4. Concluding remarks

The results from this short note show that we may process for either the MH estimator or the PMLE for estimating the common odds ratio for multiple 2×2 tables, even for small to moderate sample sizes. Since the MH estimator is easy to calculate, we may opt for that without sacrificing significant efficiency. Here both PMLE and the MH estimator converge to the true value of the common odds ratio, which is constant across the tables. Hence both the estimators converge to the theoretical value of the common odds ratio in probability also. It is also trivial to prove the convergence of $\widehat{\psi}_{\text{MH}}$ to $\widehat{\psi}_{\text{PMLE}}$ in probability by triangle inequality.

Acknowledgment

The authors wish to thank the referee for careful assessment and constructive suggestion to the paper.

Appendix

Proof of Theorem 1. Asymptotic variance of ψ_{MH} is given by

$$\text{Var}(\widehat{\psi}_{\text{MH}}) = \psi^2 \frac{\sum_{i=1}^{k} \mu_i^2/\omega_i}{\left(\sum_{i=1}^{k} \mu_i\right)^2} = \psi^2 f_\omega(\mu), \quad \text{say},$$

where $\mu_i = \left(\left(\frac{n_{Ai} n_{Bi}}{N_i}\right) q_{Ai} p_{Bi}\right)$ and $\psi^2/\omega_i = \text{Var}(\widehat{\psi}_i)$; and the asymptotic variance of the profile likelihood based estimator is

$$\text{Var}(\widehat{\psi}_{\text{PMLE}}) = \psi^2 / \sum_{i=1}^{k} \omega_i.$$

See Hauck (1979, p. 818) and Gart (1992). Note that $\mathrm{Var}(\widehat{\psi}_{\mathrm{PMLE}}) = \psi^2 f_\omega(\omega)$. Taylor series expansion of $f_\omega(\mu)$ around ω gives

$$f_\omega(\omega) + (\mu-\omega)'\beta_1 + (\mu-\omega)'\beta_2(\mu-\omega)/2 + \cdots,$$

where $\beta_1 = \frac{\partial}{\partial\mu} f_\omega(\mu)|_{\mu=\omega}$ and $\beta_2 = \frac{\partial^2}{\partial\mu_i\partial\mu_j} f_\omega(\mu)|_{\mu=\omega}$. It is easy to see that $\beta_{1i} = 0$ implying that $(\mu-\omega)'\beta_1 = 0$. Again $(\beta_2)_{ii} = \frac{2}{\omega_i\left(\sum_{i=1}^k \omega_i\right)^2} - \frac{2}{\left(\sum_{i=1}^k \omega_i\right)^3}$ and $(\beta_2)_{ij} = \frac{2}{\left(\sum_{i=1}^k \omega_i\right)^3}$ for all $i \neq j$. Now for a pre-specified k groups, $\sum_{i=1}^k \mu_i$ is constant and each μ_i^2 is convex. So their linear combination is also convex, and hence $f_\omega(\mu)$ is convex. This implies that β_2 is a positive definite matrix. If $\{\lambda_i\}_{i=1}^k$ are the eigenvalues of β_2, $\lambda_i > 0$ for all $i = 1, \ldots, k$. Now

$$\begin{aligned}
(\mu-\omega)'\beta_2(\mu-\omega) &= \frac{(\mu-\omega)'\beta_2(\mu-\omega)}{(\mu-\omega)'(\mu-\omega)}\|\mu-\omega\|^2 \\
&\leq \sup \frac{(\mu-\omega)'\beta_2(\mu-\omega)}{(\mu-\omega)'(\mu-\omega)}\|\mu-\omega\|^2 = \max_{1\leq i\leq k}\{\lambda_i\}\,\|\mu-\omega\|^2 \\
&\leq \left(\sum_{i=1}^k \lambda_i\right)\|\mu-\omega\|^2 = \mathit{trace}(\beta_2)\,\|\mu-\omega\|^2 = \left\{\frac{2\left(\sum_{i=1}^k 1/\omega_i\right)}{\left(\sum_{i=1}^k \omega_i\right)^2} - \frac{2k}{\left(\sum_{i=1}^k \omega_i\right)^3}\right\}\|\mu-\omega\|^2 \\
&\leq \left\{\frac{2\left(\sum_{i=1}^k 1/\omega_i\right)}{\left(\sum_{i=1}^k \omega_i\right)^2}\right\}\|\mu-\omega\|^2 \leq \left\{\frac{2\left(\sum_{i=1}^k 1/\omega_i\right)^3}{k^4}\right\}\|\mu-\omega\|^2 = O\left((kn)^{-(3+\delta)}\right),
\end{aligned}$$

for some $\delta > 0$, since $\left(\frac{1}{k}\sum_{i=1}^k 1/\omega_i\right) \geq \frac{k}{\sum_{i=1}^k \omega_i}$. Hence the result follows. $\square$

References

Anscombe, F.J., 1956. On estimating bivariate response relations. Biometrika 43, 461–464.
Biswas, A., Hwang, J.-S., 2011. Distribution of odds ratio in 2 × 2 contingency tables: adjustment for correlation. Journal of Biopharmaceutical Statistics 21, 136–150.
Bohning, D., Kuhnert, R., Rattanasiri, S., 2008. Meta-Analysis for Binary Data Using Profile Likelihood. Chapman & Hall/CRC Press, Boca Raton.
Breslow, N., 1981. Odds ratio estimators when the data are sparse. Biometrika 68, 73–84.
Breslow, N.E., Liang, K.-Y., 1982. The variance of the Mantel–Haenszel estimator. Biometrics 38, 943–952.
Falanders, W.D., 1985. A new variance estimator for Mantel–Haenszel odds ratio. Biometrics 41, 637–642.
Gart, J.J., 1992. On the combination of relative risks. Biometrics 18, 601–610.
Hauck, W.W., 1979. The large sample variance of the Mantel–Haenszel estimator of a common odds ratio. Biometrics 35, 817–819.
Mantel, N., Haenszel, W., 1959. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22, 719–748.
Murphy, S.A., van der Vaart, A.W., 2000. On profile likelihood. Journal of the American Statistical Association 95, 449–465.
Phillps, A., Holland, P.W., 1987. Estimators of variance of the Mantel–Haenszel log-odds-ratio estimate. Biometrics 43, 425–431.
Pigeot, I., 1990. A class of asymptotically efficient non-iterative estimators of a common odds ratio. Biometrika 77, 420–423.
Pigeot, I., 1991. A simulation study of estimators of common odds ratio in several 2 × 2 tables. Journal of Statistical Computation and Simulation 38, 65–82.
Robins, J., Breslow, N., Greenland, S., 1986. Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. Biometrics 42, 311–323.
Tarone, R.E., Gart, J.J., Hauck, W.W., 1983. On the asymptotic inefficiency of certain non iterative estimators of a common relative risk or odds ratio. Biometrica 70, 519–522.