ON TESTS FOR HOMOGENEITY OF ODDS RATIO

Tosiya SATO

Department of Epidemiology, School of Health Sciences

University of Tokyo, Tokyo 113, Japan

SUMMARY

Tests for homogeneity of the odds ratio for a series of 2×2 tables when the strata are large are examined by Monte Carlo experiments. The conditional score test and a test based on the Mantel-Haenszel estimator with the conditional variance are shown to perform better. On the other hand, the unconditional score test, though Liang and Self (1985, *Biometrika* 72, 353-358) reported that its large-sample chi-squared approximation was adequate when there were as many as ten tables with as few as ten observations per table, cannot be recommended from our results.

1. はじめに

疾病の後向き研究である case-control 研究では通常データは 2×2 分割表にまとめられ、疾病と曝露との関連の程度を表すオッズ比が最も興味のあるパラメータとなる。交絡要因が存在する場合には関連の程度が歪められたり、極端な場合には関連の方向が歪められる恐れがあるので、交絡から生じる偏った比較を避けるためには交絡要因のレベルに応じて層別にサンプリングを行う、あるいは交絡変数により事後層別を行ってそれぞれのサブグループ内で比較を行う必要がある。複数の 2×2 分割表を解析する場合、もし各サブグループで2次の交互作用がないときには（つまり、control の曝露割合がサブグループごとに変化しても、オッズ比は共通な場合）、共通オッズ比に関する推測を行うことが正当化される。

各層別のオッズ比が均一であるかどうかはどうすれば調べることができるであろうか。この仮説検定は logistic モデルや log-linear モデルの2次の交互作用の検定と同一であるが、 2×2 分割表の場合には共通オッズ比の推定量と密接な関わりがあるので、本論文ではオッズ比の要約推定量に関連づけていくつかの均一性検定を述べる。また、これらの検定は大標本のもとでの近似検定なので、パラメータのいくつかの組合せについて Monte Carlo 実験により検定の実質上の有意水準を比較する。

Key words: Breslow-Day test; Conditional score test; Tarone-Yanagawa test; Tests based on the Mantel-Haenszel estimator; Unconditional score test.

2. これまでに提案されている検定

データはK個のサブグループに層別されており、各層ごとに (x, y) は独立に二項分布 $Bi(n, p_1)$, $Bi(m, p_0)$ に従うと仮定する（層を示す添字は省略する）. Case-control 研究では (x, y) は case グループ n 人, control グループ m 人中のリスク要因の曝露を受けた数を表す. 本論文で述べる検定は「層の数はK個で固定されていて、各層の標本数 (n, m) が無限に増加する」という大標本モデルにおける近似検定である.

興味のあるパラメータはオッズ比 ϕ であるから、各層のデータの確率モデルとして二項分布モデルを考えると、 p_0 は局外パラメータとなる（ p_0 の大小は疾病-曝露間の関連の程度に影響しない）. 従って、オッズ比に関する局外パラメータ p_0 の影響を受けない推測を行うには、 $t = x + y$ を既知とした条件つき分布である非心超幾何分布を用いることが望ましい. ここで、 $f(x; \phi)$ を非心超幾何分布の確率とし、また、

$$S(\phi) = \frac{d \log f(x; \phi)}{d \phi}, \quad I(\phi) = E \left[- \frac{d^2 \log f(x; \phi)}{d \phi^2} \right],$$

とすると、条件つき尤度から次の条件つきスコア検定、

$$T_1(\hat{\phi}_e) = \Sigma [S(\hat{\phi}_e)^2 / I(\hat{\phi}_e)],$$

を得る（特に断わらない限り、和は全ての層にわたってとるものとする）. $\hat{\phi}_e$ は共通オッズ比の条件つき最尤推定量である. 簡単な計算から、

$$T_1(\hat{\phi}_e) = \Sigma \frac{[x - E(x; \hat{\phi}_e)]^2}{\text{var}(x; \hat{\phi}_e)}, \quad (1)$$

となることが分かる. ただし、 $E(x; \phi)$, $\text{var}(x; \phi)$ は ϕ が与えられたもとでの非心超幾何分布の正確な期待値と分散

$$E(x; \phi) = \Sigma_u u f(u; \phi), \\ \text{var}(x; \phi) = \Sigma_u [u - E(x; \phi)]^2 f(u; \phi),$$

（ $u \in [\max(0, t - m), \min(n, t)]$ ）である. T_1 はオッズ比が均一であるという帰無仮説のもとで漸近的に自由度 $K - 1$ の χ^2 分布に従う.

条件つき分布を考えるかわりに、二項分布モデルのもとでの全尤度から（無条件の）スコア検定、

$$T_2(\hat{\phi}_u) = \Sigma \frac{[x - E^A(x; \hat{\phi}_u)]^2}{\text{var}^A(x; \hat{\phi}_u)}, \quad (2)$$

が得られる（付録参照）．ただし $\hat{\phi}_u$ は共通オッズ比の最尤推定量， $E^A(x; \phi) = E^A$ は ϕ が与えられたもとの漸近的な期待値で次の 2 次式の適切な根，

$$E^A(m - t + E^A) = \phi(n - E^A)(t - E^A), \quad (3)$$

および $\text{var}^A(x; \phi)$ は漸近分散，

$$\text{var}^A(x; \phi) = \left(\frac{1}{E^A} + \frac{1}{n - E^A} + \frac{1}{t - E^A} + \frac{1}{m - t + E^A} \right)^{-1},$$

である． T_2 は帰無仮説のもとで漸的に自由度 $K - 1$ の χ^2 分布に従う．この検定は最初に Zelen (1971) によって非心超幾何分布の正規近似から導かれたがその導出に誤りがあり，後に Zelen の誤りを Halperin et al. (1977) が訂正し (2) 式を与えている．

スコア検定中の最尤推定量 $\hat{\phi}_u$ を ϕ の任意の一致推定量で置き換えても，標本数が十分大きければ均一性の仮説のもとで近似的に自由度 $K - 1$ の χ^2 分布に従うと考えてもいいだろう．そこで， ϕ の一致推定量として Mantel and Haenszel (1959) が提案した，

$$\hat{\phi}_{MH} = \frac{\Sigma x(m - y)/N}{\Sigma y(n - x)/N} = \frac{\Sigma R}{\Sigma S},$$

（ここで， $R = x(m - y)/N$ ， $S = y(n - x)/N$ ， $N = n + m$ ）を T_2 に代入した，

$$T_3(\hat{\phi}_{MH}) = \Sigma \frac{[x - E^A(x; \hat{\phi}_{MH})]^2}{\text{var}^A(x; \hat{\phi}_{MH})}, \quad (4)$$

を Breslow and Day (1980) が提案した (Breslow-Day 検定) ．

しかし，実際には T_3 は均一性の仮定のもとで自由度 $K - 1$ の χ^2 分布に従う確率変数よりも確率的に大きくなる (Tarone, 1985)．このため，一致推定量ではあるが漸的に有効ではない Mantel-Haenszel 推定量を one-step 改良して最尤推定量に近づけ一次の有効性を持たせた，

$$\hat{\phi}_{MH}^1 = \hat{\phi}_{MH} + \hat{\phi}_{MH} \frac{\Sigma [x - E^A(x; \hat{\phi}_{MH})]}{\Sigma \text{var}^A(x; \hat{\phi}_{MH})},$$

を T_3 に代入して Breslow-Day 検定を修正した,

$$T_4(\hat{\phi}_{MH}^{-1}) = T_3(\hat{\phi}_{MH}) - \frac{\{\Sigma [x - E^A(x; \hat{\phi}_{MH})]\}^2}{\Sigma \text{var}^A(x; \hat{\phi}_{MH})}, \quad (5)$$

を用いるべきであることを Tarone (1985) と柳川 (1986) が提案した (Tarone-柳川検定). T_4 は帰無仮説のもとで漸近的に自由度 $K-1$ の χ^2 分布に従う.

3. Mantel-Haenszel 推定量にもとづく検定

Breslow-Day 検定も Tarone-柳川検定も $\hat{\phi}_{MH}$ にもとづいて odds ratio の均一性の推論を行っているわけではなく, $\hat{\phi}_{MH}$ を最尤推定量の近似として検定を構成している. ここでは最尤推定量の近似としてではなく, 直接 Mantel-Haenszel 推定量にもとづいた均一性検定を考えてみよう. Mantel-Haenszel 推定量は次の推定関数,

$$\Sigma (R - \phi S) = 0,$$

の解であるため (Davis, 1985), Mantel-Haenszel 推定量にもとづく推測は $R - \phi S$ を用いて行うべきである. 条件つき分布のもとでも無条件分布のもとでも $E(R - \phi S) = 0$ であり, 各層の $R - \phi S$ は仮定した大標本モデルと適当な正則条件のもとで漸近的に正規分布に従う. 母共通オッズ比 ϕ を一致推定量である $\hat{\phi}_{MH}$ で置き換えた各層の $R - \hat{\phi}_{MH} S$ は均一性の仮定からのズレを表すので, 全体としてのズレは,

$$\Sigma \frac{(R - \hat{\phi}_{MH} S)^2}{\text{var}(R - \hat{\phi}_{MH} S)},$$

が漸近的に自由度 $K-1$ の χ^2 分布に従うことから調べることができる. $\text{var}(R - \hat{\phi}_{MH} S)$ の推定量はいくつかのものが提案されており, それによって多少異なった検定統計量が導かれる.

まず, Breslow (1981) が提案した条件つき分散,

$$\text{var}_c(R - \hat{\phi}_{MH} S) = N^{-2} \Sigma_u [u(m-t+u) - \hat{\phi}_{MH}(n-u)(t-u)]^2 f(u; \hat{\phi}_{MH})$$

を用いると,

$$T_5(\hat{\phi}_{MH}) = \Sigma \frac{(R - \hat{\phi}_{MH} S)^2}{\text{var}_c(R - \hat{\phi}_{MH} S)}, \quad (6)$$

が得られる.

また, Hauck (1979) は $R - \hat{\phi}_{MH} S$ の漸近分散,

$$\text{var}^A(R - \hat{\phi}_{MH} S) = \hat{\phi}_{MH}^2 S^2 \left(\frac{1}{x} + \frac{1}{n-x} + \frac{1}{y} + \frac{1}{m-y} \right),$$

を導いた. この推定量は行または列の交換に対し不変性を持たない, すなわち $\text{var}^A(S - R/\hat{\phi}_{MH}) \neq \text{var}^A(R - \hat{\phi}_{MH} S)/\hat{\phi}_{MH}^2$, また (x, y) の変動の影響を直接こうむるため, Gart は $\text{var}^A(R - \hat{\phi}_{MH} S)$ 中の (x, y) を $\hat{\phi}_{MH}$ を既知としたときの (3) 式の漸近期待値で置き換えることを提案した (Ury, 1982). この Hauck-Gart 推定量を用いて,

$$T_6(\hat{\phi}_{MH}) = \Sigma \frac{(R - \hat{\phi}_{MH} S)^2}{\text{var}_{HG}(R - \hat{\phi}_{MH} S)}, \quad (7)$$

が得られる.

行または列の交換に対して不変である $\text{var}(R - \hat{\phi}_{MH} S)$ の推定量としては,

$$\text{var}_S(R - \hat{\phi}_{MH} S) = \hat{\phi}_{MH} [(Q+1/N)R + (P+1/N)S],$$

ただし $P = (x+m-y)/N$, $Q = (y+n-x)/N$, が提案されている (Sato, 1988). しかし, この推定量も (x, y) の変動の影響を受けるので (x, y) に $\hat{\phi}_{MH}$ 既知としたときの漸近期待値を代入して (7) 式と同様に Gart の修正を行った Sato-Gart 推定量を用いて,

$$T_7(\hat{\phi}_{MH}) = \Sigma \frac{(R - \hat{\phi}_{MH} S)^2}{\text{var}_{SG}(R - \hat{\phi}_{MH} S)}, \quad (8)$$

を得る.

4. Monte Carlo 実験のデザイン

前節までに示した T_1 から T_7 まで 7 つの検定統計量について, 実質的な有意水準を調べるために Monte Carlo 実験を行った. パラメータ K , ϕ , p_1 は疫学研究に現われるような範囲で選んだ. Case の曝露+の比率 p_1 は 0.3 から 0.7 の範囲になるように,

$$p_{1k} = 0.4 + 0.3 \times (k-1)/(K-1), \quad k = 1, 2, \dots, K$$

とし, 層の数は $K = 5, 10, 20$, オッズ比は $\phi = 1.0, 3.5$ とした. (n, m) の値は, 特に中程度の標本数でのパフォーマンスを問題にしているので, $(10, 10)$, $(20, 20)$, $(40, 40)$ の 3 通りを選び, 総標本数は 100 から 1600 の範囲となった. 検定の名義的な有意水準は $\alpha = 0.05$ とした. 合計

18の実験で各2000回の trial を実行した。これより、各検定が名義的な有意水準どおりならば、実質的な有意水準の95%信頼区間は(0.04, 0.06)となる。

二項乱数はSASの組み込み関数RANBINを用いて発生させた。計算は東京大学大型計算機センターHITAC M682H/S820システム上で実行した。

5. 結果および考察

$\phi = 1.0$ の結果をTable 1に、 $\phi = 3.5$ の結果をTable 2に示す。どちらの場合も、条件つきスコア検定 T_1 と条件つき分散を用いた $\hat{\phi}_{MH}$ にもとづく検定 T_5 は名義的な水準と非常に近い値を示している。また、スコア検定 T_2 、Breslow-Day 検定 T_3 、Tarone-柳川 検定 T_4 、Hauck-Gart 推定量を用いた $\hat{\phi}_{MH}$ に基づく検定 T_6 は radical な方向にバイアスがある。 $\phi = 1.0$ で $n=m=20$ 以上のとき、これら4つの検定のパフォーマンスは等しい。これは、 $\phi = 1$ のとき $\hat{\phi}_{MH}$ は漸近有効推定量であり、(無条件の)最尤推定量のバイアスもほとんどない(Hauck, Anderson and Leahy, 1982)ためであろう。 $n=m=10$ のときは T_3 と T_4 の結果が等しく、 T_2 と T_6 よりもわずかながら名義的な有意水準に近い。この傾向は $\phi = 3.5$ の場合により明確に現われている。また、このときにはやはりわずかな差ではあるが T_3 よりも T_4 が名義的な水準に近かった。ただし、 ϕ の値にかかわらず、 $n=m=10$ 、 $K=20$ では4つの検定ともに非常に radical な結果であった。これらの結果とは逆に Sato-Gart 推定量を用いた $\hat{\phi}_{MH}$ にもとづく検定 T_7 は conservative なバイアスがある。特に $n=m=10$ 、 $K=20$ では非常に conservative である。

条件つきスコア検定は、ここで検討した全てのデザインにわたって有効であった。条件つき最尤推定量にもとづいて共通オッズ比の推測を行う場合には、均一性検定も条件つきスコア検定を用いるべきであろう。

一方、最尤推定量にもとづいて共通オッズ比の推測を行う場合には、中程度の標本数でもバイアスがあることが知られている(Hauck et al., 1982)。均一性検定に関しては、Liang and Self (1985)がシミュレーションの結果から「少なくとも $n=m=10$ で多くとも $K=10$ であればスコア検定の χ^2 近似は十分である」と述べているが、本研究の他のデザインも考えると radical なバイアスがあることは明らかである。むしろ、Breslow-Day 検定さらには Tarone-柳川検定の方がバイアスが小さく、均一性検定にはこちらを用いた方がいい。Breslow-Day 検定と Tarone-柳川検定はスコア検定の近似であり、特に Breslow-Day 検定は計算の簡便さのみから提案された(Breslow and Day, 1980)にもかかわらず、なぜスコア検定よりも名義的な水準に近いのかは不明であり、今後検討を必要とする。

Mantel-Haenszel 推定量にもとづいた推測を行う場合には、均一性検定として条件つき分散を用いた $\hat{\phi}_{MH}$ にもとづく検定を行えば、かなり広い範囲のデザインにわたって満足のいく結果が得られる。しかし、条件つき分散の計算は条件つきスコア検定同様かなり大変であり、Mantel-Haenszel 推定量の最大の特徴である計算の簡便さが損なわれてしまう。従って、この場合にも近似検定としては Tarone-柳川検定を用いるか、あるいは多少保守的な結果がほしい場合には Sato-Gart 推定量を用いた $\hat{\phi}_{MH}$ にもとづく検定を行えばよいだろう。

Table 1. Monte Carlo significance levels of the homogeneity tests,
in percent, $\phi=1.0$

n	m	K	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇
10	10	5	5.0	6.0	5.8	5.8	4.8	6.1	3.8
		10	4.8	5.8	5.8	5.8	4.8	6.3	3.8
		20	5.0	7.0	6.8	6.8	5.1	7.2	3.3
20	20	5	5.3	5.5	5.5	5.5	5.0	5.5	4.8
		10	4.6	5.1	5.1	5.1	4.5	5.1	3.8
		20	4.8	6.0	6.0	6.0	4.8	6.1	4.0
40	40	5	5.4	5.6	5.6	5.6	5.3	5.6	5.1
		10	4.3	4.8	4.8	4.8	4.3	4.8	4.1
		20	4.6	5.1	5.1	5.1	4.6	5.0	4.1

Table 2. Monte Carlo significance levels of the homogeneity tests,
in percent, $\phi=3.5$

n	m	K	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇
10	10	5	4.7	6.1	5.4	5.3	4.7	6.1	3.8
		10	4.6	5.6	5.3	5.1	4.5	5.6	3.6
		20	4.5	7.0	6.3	5.8	4.6	7.5	3.1
20	20	5	5.3	5.6	5.4	5.3	5.0	5.5	4.3
		10	4.1	4.8	4.6	4.6	4.8	5.5	4.0
		20	3.7	5.5	5.0	5.0	4.6	5.6	3.4
40	40	5	4.8	5.3	5.3	5.1	5.0	5.2	4.8
		10	5.1	5.6	5.5	5.5	5.3	5.7	4.8
		20	4.6	5.6	5.5	5.4	5.1	5.7	4.5

文 献

- Breslow, N.E. and Day, N.E. (1980). Statistical Methods in Cancer Research: Vol.1 The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer.
- Breslow, N.E. (1981). Odds ratio estimators when the data are sparse. *Biometrika* 68, 73-84.
- Cornfield, J. (1956). A statistical problem arising from retrospective studies. *Proceedings of the Third Berkeley Symposium* 4, 135-148.
- Davis, L.J. (1985). Generalization of the Mantel-Haenszel estimator to nonconstant odds ratios. *Biometrics* 41, 487-495.
- Halperin, M., Ware, J.H., Byar, D.P., Mantel, N., Brown, C.C., Koziol, J., Gail, M. and Green, S.B. (1977). Testing for interaction in an $I \times J \times K$ contingency table. *Biometrika* 64, 271-275.
- Hauck, W.W. (1979). The large sample variance of the Mantel-Haenszel estimator of the common odds ratio. *Biometrics* 35, 817-819.
- Hauck, W.W., Anderson, S. and Leahy, F.J. III. (1982). Finite-sample properties of some old and some new estimators of a common odds ratio from multiple 2×2 tables. *Journal of the American Statistical Association* 77, 145-152.
- Liang, K.-Y. and Self, S.G. (1985). Tests for homogeneity of odds ratio when the data are sparse. *Biometrika* 72, 353-358.
- 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.
- Sato, T. (1988). Confidence limits to the common odds ratio based on the Mantel-Haenszel estimator. Submitted for publication.
- Tarone, R.E. (1985). On heterogeneity tests based on efficient scores. *Biometrika* 72, 91-95.
- Ury, H.K. (1982). Hauck's approximate large-sample variance of the Mantel-Haenszel estimator. *Biometrics* 38, 1094-1095.
- 柳川 堯 (1986). 離散多変量データの解析 東京 共立出版.
- Zelen, M. (1971). The analysis of several 2×2 contingency tables. *Biometrika* 58, 129-137.

付 録

ここでは無条件のスコア検定が(2)式と一致することを示す。今(x, y)の同時確率を形式的に $t = x + y$ において(tを固定するわけではない) ϕ と p_o を用いて書き直すと,

$$P = \binom{n}{x} \binom{m}{t-x} \phi^x \frac{p_o^t (1-p_o)^{N-t}}{[\phi p_o + (1-p_o)]^n} \quad (A.1)$$

を得る。log P を (ϕ , p_o) で偏微分すると,

$$S(\phi) = \frac{\partial}{\partial \phi} \log P = \frac{x}{\phi} - \frac{n p_o}{\phi p_o + (1-p_o)} \quad (A.2)$$

$$\frac{\partial}{\partial p_o} \log P = - \frac{n(\phi-1)}{\phi p_o + (1-p_o)} + \frac{t}{p_o} - \frac{N-t}{1-p_o} \quad (A.3)$$

であり, $p_1 = \phi p_o / [\phi p_o + (1-p_o)]$ であるから $n p_1 = E^A$ とおくと, (A.2) 式は,

$$S(\phi) = (x - E^A) / \phi.$$

また同様に E^A を用いると ϕ 既知のとき (A.3) 式をゼロにする p_o は,

$$E^A(m - t + E^A) = \phi(n - E^A)(t - E^A)$$

を満たし, (3) 式が得られた.

さらに Fisher 情報行列 I は,

$$\begin{aligned} I &= \begin{pmatrix} E\left[-\frac{\partial^2}{\partial \phi^2} \log P\right] & E\left[-\frac{\partial^2}{\partial \phi \partial p_o} \log P\right] \\ E\left[-\frac{\partial^2}{\partial p_o \partial \phi} \log P\right] & E\left[-\frac{\partial^2}{\partial p_o^2} \log P\right] \end{pmatrix} \\ &= \frac{1}{n \phi^2} \begin{pmatrix} A & B \\ B & B(B + n m \phi) / A \end{pmatrix} \end{aligned}$$

(ただし $A = E^A(n - E^A)$, $B = [E^A + \phi(n - E^A)]^2$) となって, その逆行列の第(1,1)要素を $I(\phi)^{-1}$ とすると, $E^A(m - t + E^A) = \phi(n - E^A)(t - E^A)$ に注意して,

$$I(\phi)^{-1} = \frac{\phi[E^A + \phi(n - E^A)]^2 + nm\phi^2}{mE^A(n - E^A)}$$

$$= \phi^2 \left(\frac{1}{E^A} + \frac{1}{n - E^A} + \frac{1}{t - E^A} + \frac{1}{m - t + E^A} \right) \quad (\text{A. 4})$$

が導かれる。従って、共通オッズ比の最尤推定量 $\hat{\phi}_u$ が与えられたときのスコア検定は、

$$\Sigma S(\hat{\phi}_u)^2 I(\hat{\phi}_u)^{-1}$$

$$= \Sigma \frac{[x - E^A(x; \hat{\phi}_u)]^2}{\text{var}^A(x; \hat{\phi}_u)} \quad (\text{A. 5})$$

が均一性の仮定のもとで漸近的に自由度 $K - 1$ の χ^2 分布に従うことから構成できる。(A. 5)式は(2)式にはかならない。

また、単一の 2×2 分割表でオッズ比の近似信頼区間を推定する Cornfield (1956)の方法は非心超幾何分布の正規近似から導かれたが、この結果より無条件のスコア検定として解釈することもできる。
[Received May, 1988]