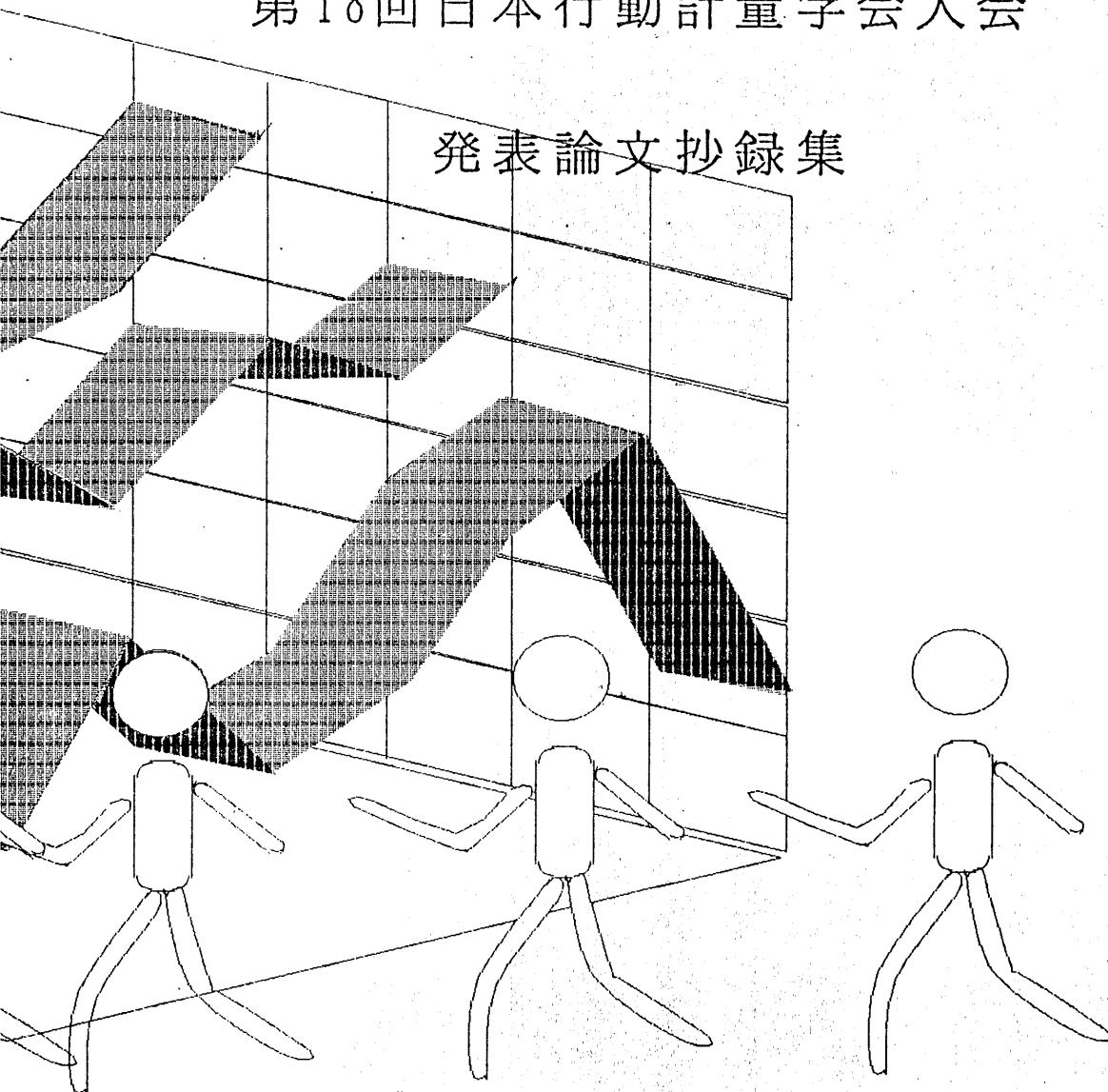


第18回日本行動計量学会大会

発表論文抄録集



平成2年9月19日(水)～21日(金)
東京女子大学現代文化学部
〒181東京都三鷹市牟礼4-3-1
京王帝都井の頭線「井の頭公園」駅下車

層別四分表の共通オッズ比の正確な信頼区間 計算のためのアルゴリズム

高木 廣文 (統計数理研究所)

1. はじめに

疫学研究等である疾患に対する特定の要因の有無による影響を検討する場合、交絡因子の影響を除くために標本(患者, 対照)を層別し, K 個の四分表に共通するオッズ比を求めることがよく行なわれる. 通常, 共通オッズ比は Mantel-Haenszel(MH) 推定量を用い, その信頼区間は MH 推定量の漸近分散 (Robins et al, 1986) を用いて算出するか, 検定統計量 (Sato, 1990) を基に算出すれば近似はよいとされている.

一方, 条件付尤度に基づき, 非心超幾何分布から推定を行なう「正確な」方法がある (Gart, 1970, 1971, Zelen, 1971). 共通オッズ比 ψ は非心超幾何分布の非心度となるが, 条件付最尤推定量の計算には, 組合せ計算を含むため, 標本数が比較的小さい場合にしか実用的ではなかった. しかし, 現在では比較的大標本であっても計算が可能になった (Satten and Kupper, 1990, and Takagi, 1990).

共通オッズ比の信頼区間を非心超幾何分布に基づいて推定する場合, K 個の四分表の患者数が全て n となると, 最悪のアルゴリズムでは n^K のオーダーの計算量が必要になる. Mehta ら (1985) は Network Algorithm を提唱し, それにより計算量が $(nK)^2$ のオーダーになることを示した. しかし, Mehta らの方法は, 単純な再帰的なアルゴリズムと同等であり, Network Algorithm は必ずしも必要ではない.

今回は, 層別四分表の共通オッズ比の信頼区間の推定を行うためのアルゴリズムを報告する.

2. 条件付分布と正確な信頼区間

K 個の四分表のうち第 k 層の四分表の患者数, 対照数を n_k, m_k とし, 要因 X があるものを各 x_k, y_k とする. いま, $t_k = x_k + y_k$ が与えられているものとする. $l_k = \max(0, t_k - m_k)$, $u_k = \min(n_k, t_k)$ とし, $L = \sum_k l_k$, $U = \sum_k u_k$ とする.

K 個の四分表の周辺度数が固定されている場合, x_k の和が s となる条件付分布は,

$$g(s|\psi) = C_s \psi^s / \sum_{r=L}^U C_r \psi^r, \quad L \leq s \leq U, \quad (1)$$

により与えられる (Zelen, 1971). ここで,

$$C_s = \sum_{R(s)} \prod_{k=1}^K \binom{n_k}{r_k} \binom{m_k}{t_k - r_k}, \quad (2)$$

であり, $R(s)$ は $\sum_k r_k = s$ となる K 個の四分表の組合せの全集合である.

正確な $100(1-\alpha)\%$ 信頼区間を $(\hat{\psi}_L, \hat{\psi}_U)$ とすると,

$$\sum_{s=L}^U g(s|\hat{\psi}_L) = \frac{\alpha}{2}, \quad \text{および} \quad \sum_{s=L}^U g(s|\hat{\psi}_U) = \frac{\alpha}{2}, \quad (3)$$

を満たすような $\hat{\psi}_L$ と $\hat{\psi}_U$ を求めればよい. ただし, $s = L$ の場合, $\hat{\psi}_L = 0$ とし, $s = U$ の場合, $\hat{\psi}_U = \infty$ とする.

問題は, (2) 式の係数 C_s を求めるための計算が組合せを含むため, 煩雑であることと数値オーバーフローの対策が必要なことである. 係数 C_s さえ求められれば, 共通オッズ比の信頼区間の推定は, 簡単な反復計算により可能である.

3. 再帰的アルゴリズム

まず、各層ごとに、

$$W(k, r_k) = \binom{n_k}{r_k} \binom{m_k}{l_k - r_k}, \quad l_k \leq r_k \leq u_k, \quad k = 1, \dots, K,$$

を求めておく。(2)式の各係数を求めるには、

[1] $C(1, r_1) = W(1, r_1)$, $l_1 \leq r_1 \leq u_1$, とおく。

[2] $L_k = \sum_{j=1}^k l_j$ および $U_k = \sum_{j=1}^k u_j$ とする。

$$C(k+1, z_{k+1}) = \sum_{z_k + r_{k+1} = z_{k+1}} C(k, z_k) \cdot W(k+1, r_{k+1}), \quad (L_k \leq z_k \leq U_k, l_{k+1} \leq r_{k+1} \leq u_{k+1}),$$

を求める。

[3][2] のステップを繰り返すことにより、 $C_z = C(K, z_K)$ が求められる。

上記の手順 [2] による最大の計算量は、 $n = \max\{n_1, \dots, n_K\}$ とすると、

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

となり、計算量が $(nK)^2$ のオーダーになることが分かる。

ところで、実際の計算を行なう場合、上記の手順では計算途中でオーバーフローを起こす可能性が高い。そこで、 $lW(k, r_k) = \ln W(k, r_k)$ および $lC(k, z_k) = \ln C(k, z_k)$ のように対数をとる。

$$lC(k+1, z_{k+1}) = \ln \left[\sum_{z_k + r_{k+1} = z_{k+1}} \exp[lC(k, z_k) + lW(k+1, r_{k+1}) - \Delta(z_{k+1})] \right] + \Delta(z_{k+1})$$

という関係を用いればよい。ここで、 $\Delta(z_{k+1}) = \max[lC(k, z_k) + lW(k+1, r_{k+1}); z_{k+1} = z_k + r_{k+1}, L_k \leq z_k \leq U_k, l_{k+1} \leq r_{k+1} \leq u_{k+1}]$ とする。実際の計算では、 $\Delta(z_{k+1})$ を適当な配列に記憶させ、計算の繰り返し毎に値を比較して更新していけばよい。

ところで、(3)式の計算を行なう場合、 $C_r \psi^r$ という計算が必要であり、そのままでは数値オーバーフローの原因になり得る。この問題を避けるためには、 $\hat{\psi}_L$ については、

$$\sum_{z=L}^U \exp[\ln C_z + z \ln \hat{\psi}_L - \Delta] / \sum_{r=L}^U \exp[\ln C_r + r \ln \hat{\psi}_L - \Delta]$$

という計算を行なえばよい。ここで、 $\Delta = \max[\ln C_r + r \ln \hat{\psi}_L; L \leq r \leq U]$ とする。 $\hat{\psi}_U$ についても同様である。上記の計算では、 C_r ではなく $\ln C_r$ を求めておけばよいことが分かる。

[参考文献]

- Gart, J.J. (1970). *Biometrika* 57, 471-475.
 Gart, J.J. (1971). *Review of the International Statistical Institute* 39, 148-169.
 Mantel, N. and Haenszel, W. (1959). *Journal of the National Cancer Institute* 22, 719-748.
 Mehta, C.R., Patel, N.R., and Gray, R. (1985). *Journal of the American Statistical Association* 80, 969-973.
 Robins, J.M., Breslow, N.E., and Greenland, S. (1986). *Biometrics* 42, 311-323.
 Sato, T. (1990). *Biometrics* 46, 71-80.
 Satten, G. A. and Kupper L.L. (1990). *Biometrics* 46, 217-223.
 Takagi, H. (1990). *Computer Science Monographs*, submitted.
 Zelen, M. (1971). *Biometrika* 58, 129-137.