Local Odds Ratio Estimation for Stratified Contingency Tables with Multiple Responses

Ivy Liu[1] and Thomas Suesse[2]

[1] *School of Mathematics, Statistics and Operations Research*
Victoria University of Wellington
Wellington, New Zealand

[2] *School of Mathematics and Applied Statistics*
University of Wollongong
Wollongong, Australia

The International Biometric Society Australasian Region Conference
December 2009

1. Local Odds Ratios

   — Highly Stratified Contingency Tables with Multiple Responses

2. Maximum Likelihood (ML) Approach

3. Mantel–Haenszel (MH) Approach

4. Example

5. Simulation Results

I. Local Odds Ratios

– For an ordinary $2 \times c$ contingency table,

Group	Responses					
	1	2	3	$\cdots$	$c-1$	c
1	$\pi_{1\|1}$	$\pi_{2\|1}$	$\pi_{3\|1}$		$\pi_{c-1\|1}$	$\pi_{c\|1}$
2	$\pi_{1\|2}$	$\pi_{2\|2}$	$\pi_{3\|2}$		$\pi_{c-1\|2}$	$\pi_{c\|2}$

- The $\pi_{j|i}$ is the probability of response on level j for a subject in row i.
- $\sum_j \pi_{j|i} = 1$ for all $j = 1, 2$.
- The local odds ratio is defined by

$$\Psi_j = \frac{\pi_{j|1}\pi_{(j+1)|2}}{\pi_{j|2}\pi_{(j+1)|1}}, \quad j = 1, \ldots, c-1.$$

– For a stratified $2 \times c \times K$ contingency table, the kth table:

	Responses					
Group	1	2	3	$\cdots$	$c-1$	c
1	$\pi_{1\|1k}$	$\pi_{2\|1k}$	$\pi_{3\|1k}$		$\pi_{c-1\|1k}$	$\pi_{c\|1k}$
2	$\pi_{1\|2k}$	$\pi_{2\|2k}$	$\pi_{3\|2k}$		$\pi_{c-1\|2k}$	$\pi_{c\|2k}$

• The local odds ratio is defined by

$$\Psi_{j|k} = \frac{\pi_{j|1k}\pi_{(j+1)|2k}}{\pi_{j|2k}\pi_{(j+1)|1k}}, \quad j = 1, \ldots, c-1 \text{ and } k = 1, \ldots, K$$

• It shows the associations between the group and responses controlling for a possibly confounding variable.

• When the stratified table is highly sparse, we often summarise the association across strata by assuming the common odds ratios, i.e.,

$$\Psi_j = \Psi_{j|1} = \cdots = \Psi_{j|K}$$

• The cell counts in the contingency table described above are mutually exclusive, i.e., all subjects must fit into one and only one cell.

– Multiple Responses

• Non-mutually exclusive cell counts occurs when respondents may select any number out of c outcome categories.

– Example

The students of a second-year applied statistics course were asked to tick their favourite bar in Wellington. The study recorded the features of the bars (eg., "drink deals", "pool tables", and "sports TV"). Each student also answered some personal questions (such as major, gender, working status, smoking status, etc). Each bar may have more than one feature.

We are interested to find the association between working status and preferred features of the bars controlling on students' majors.

The following table shows the $2 \times 3 \times 6$ "marginal" contingency table (for 6 different majors) and the "complete" information on the multiple response profile.

						Response Profile								
						1	0	0	0	0	1	1	1	1
			items		total	2	0	0	1	1	0	0	1	1
major	work	1	2	3	students	3	0	1	0	1	0	1	0	1
BIOL														
	yes	1	0	0	1		0	0	0	0	1	0	0	0
	no	1	1	0	1		0	0	0	0	0	0	1	0
BMAR														
	yes	11	11	4	13		0	0	2	0	2	0	5	4
	no	1	1	1	1		0	0	0	0	0	0	0	1
EBIO														
	yes	7	8	4	10		1	0	2	0	1	0	2	4
	no	3	4	5	6		1	1	0	1	0	0	0	3
OPRE														
	yes	0	0	1	1		0	1	0	0	0	0	0	0
	no	0	0	0	1		1	0	0	0	0	0	0	0
PSYC														
	yes	2	2	2	3		0	1	0	0	0	0	1	1
	no	2	2	1	3		0	0	1	0	1	0	0	1
STAT														
	yes	1	1	1	1		0	0	0	0	0	0	0	1
	no	1	1	1	1		0	0	0	0	0	0	0	1

1: drink deals, 2: pool tables, 3: sports TV

2. Maximum Likelihood (ML) Approach

Assuming the common local odds ratios for all strata, the model can be expressed as:

$$\log\left(\frac{\pi_{j|1k}\pi_{(j+1)|2k}}{\pi_{j|2k}\pi_{(j+1)|1k}}\right) = \beta_j.$$

where $\beta_j = \log\Psi_j$. The model is not a standard logit model.

– The ML inference for the model requires that the cell probabilities of the complete table are estimated under the constraints imposed by the model. The observed frequency distribution from the complete table follows a multinomial distribution.

– Lang and Agresti (1994) proposed the fitting algorithm based on the Lagrange multiplier technique. (http://www.stat.uiowa.edu/∼jblang/)

– The modified algorithm was given by Bergsma (1997) and Bergsma, *et.al.* (2009). (http://www.cmm.st)

– The model can be expressed as the form of

$$C \log A\pi = X\beta$$

where π is the vector of joint probabilities from the complete table.

– The matrix A contains 0 and 1 entries in such a pattern that when applied to π it forms the relevant marginal probabilities $\{\pi_{j|ik}\}$ to which the model applies.

– The matrix C contains 0, 1 and -1 entries in such a pattern that when applied to the log marginal probabilities, it forms the log local odds ratios for the model.

– It often has convergence problems, especially for highly sparse data.

3. Mantel–Haenszel (MH) Approach

Let

$X_{j\|ik}$:	number of subjects that select item $j = 1, \ldots, c$ in row $i = 1, 2$ and stratum $k = 1, \ldots, K$ in the marginal table.
n_{ik}:	number of subjects in row i and stratum k
$N_k = n_{1k} + n_{2k}$:	total number of subjects in stratum k.

The ordinary Mantel-Haenszel estimator $\hat{\Psi}_j$ has the following form

$$\hat{\Psi}_j = \frac{\sum_{k=1}^{K} X_{j|1k} X_{(j+1)|2k} / N_k}{\sum_{k=1}^{K} X_{(j+1)|1k} X_{j|2k} / N_k}$$

For the multiple responses, we can show that

– The ordinary MH estimator $\hat{\Psi}_j$ is dually consistent,
i.e. consistent under the **large-sample** limiting model (K is bounded while the number of subjects per stratum goes to infinity)
and the **sparse-data** limiting model (K goes to infinity with sample size, but the number of subjects per stratum remains fixed).

– For a marginal table with mutually exclusive cell counts, Greenland (1989) proposed the dually consistent variance and covariance estimators for $(\hat{\Psi}_j, \forall j = 1, \ldots, c-1)$

– For multiple responses, the cell counts in the marginal table are not mutually exclusive. The dually consistency property for Greenland's estimators doesn't hold anymore.

– We proposed the new variance and covariance estimators that are dually consistent under both multiple responses and ordinary cases.

– Our estimator = (Greenland's estimator + an additional term). When the cell counts are mutually exclusive, the additional term equals 0.

– Usually, we use log odds ratio estimators, eg.

$$L_{12} = \log \hat{\Psi}_1$$
$$L_{23} = \log \hat{\Psi}_2$$
$$L_{13} = L_{12} + L_{23}$$

4. Example

We consider 3 items (bar features)

1:	drink deals
2:	pool tables
3:	sports TV

– The MH approach gives

items (j, h)	L_{jh}	standard error
items (1, 2)	0.14	0.24
items (1, 3)	0.60	0.30
items (2, 3)	0.46	0.28

– The working status has significant effects on the choice between features 1 (“drink deals”) and 3 (“sports TV”).

– The odds of choosing a favourite bar offering “drink deals” rather than “sports TV” for a student with part/full-time jobs are $\exp(0.60) = 1.83$ times the odds for a student without jobs.

– The ML approach fails to converge for this example.

– We have tried both ML and MH approaches for another example in which the large-sample limiting model seems reasonable. The ML approach converges. Both approaches give very similar estimates when data are not highly sparse.

5. Simulation Results

We conducted a simulation study to investigate the performance of the MH estimators $\log\hat{\Psi}$ and the proposed new variance and covariance estimators.

– Choose $c = 3$.

– For given Ψ_j, we fix the marginal probabilities of the first row by setting $\pi_{j|1k} = 0.50$ for all $j = 1, \ldots, c$ and find $\pi_{j|2k}$ using Ψ_j. For simplicity, we also set $\Psi = \Psi_1$, $\Psi_2 = 1$ and $N_1 = \cdots = N_K$.

– We define the pairwise dependency between items j and h in the form of an odds ratio $\theta_{jh|ik}$:

$$\theta_{jh|ik} = \frac{P(Y_j = 1, Y_h = 1|ik)P(Y_j = 0, Y_h = 0|ik)}{P(Y_j = 0, Y_h = 1|ik)P(Y_j = 1, Y_h = 0|ik)},$$

where Y_j indicate whether a subject selects item j.

– From the marginal probabilities $\{\pi_{j|ik}, j = 1, \ldots, c\}$
and the pairwise dependency odds ratios $\{\theta_{jh|ik}, j \neq h = 1, \ldots, c\}$,
we can compute the unique set of pairwise probabilities $\{\pi_{jh|ik}, j \neq h = 1, \ldots, c\}$.

– Then the 2^c joint probabilities $\{P(Y_1 = s_1, \ldots, Y_c = s_c|ik), s_j = 0, 1; j = 1, \ldots, c\}$ in the complete table can be computed from the probabilities $\{\pi_{j|ik}, j = 1, \ldots, c\}$ and $\{\pi_{jh|ik}, j \neq h = 1, \ldots, c\}$, if a feasible solution exists Lee (1993).

Simulation results for log odds ratio estimators L when the true odds ratio $\Psi = 4$ (i.e., $\log\Psi = 1.386294$)

				mean	m.s.e.
K	N_k	$\log\Psi$	θ	L_{12}, L_{13}	L_{12}, L_{13}
1	500	1.386	0	1.400, 1.399	0.0724, 0.072
1	500	1.386	1	1.394, 1.393	0.0338, 0.0341
1	500	1.386	10	1.394, 1.394	0.024, 0.0236
10	50	1.386	0	1.401, 1.401	0.0772, 0.0763
10	50	1.386	1	1.393, 1.393	0.0357, 0.0362
10	50	1.386	10	1.395, 1.396	0.0258, 0.0258
50	10	1.386	0	1.404, 1.410	0.0962, 0.096
50	10	1.386	1	1.399, 1.398	0.0454, 0.046
50	10	1.386	10	1.399, 1.400	0.0354, 0.0352
100	5	1.386	0	1.421, 1.420	0.133, 0.132
100	5	1.386	1	1.401, 1.399	0.0623, 0.0617
100	5	1.386	10	1.406, 1.405	0.0482, 0.0483

Simulation results for the variance and covariance estimators of the log odds ratio estimators:

			$\mathrm{Var}(L_{12}), \mathrm{Cov}(L_{12}, L_{13})$ Estimates		
			Empirical	New Proposed	Greenland's
K	N_k	θ	100×mean	100×mean (10000×m.s.e.)	100×mean (10000×m.s.e.)
1	500	0	7.219, 5.062	7.056, 4.936 (0.552, 0.568)	7.056, 4.936 (0.552, 0.568)
1	500	1	3.378, 2.474	3.354, 2.440 (0.111, 0.0963)	4.965, 3.251 (2.616, 0.690)
1	500	10	2.392, 1.914	2.374, 1.905 (0.0925, 0.0803)	4.964, 3.250 (6.726, 1.870)
100	5	0	13.231, 4.025	12.686, 3.855 (9.618, 0.734)	12.686, 3.855 (9.618, 0.734)
100	5	1	6.205, 3.163	6.015, 3.072 (1.474, 0.518)	8.871, 4.097 (8.859, 1.144)
100	5	10	4.781, 3.253	4.722, 3.215 (1.335, 0.767)	9.654, 5.248 (25.929, 4.608)

Note: For multiple responses, $\theta > 0$. Define $\theta = 0$ for the cases that each subjects can only select one item.

References

Bergsma, W. P. (1997). *Marginal models for categorical data*. Tiburg: Tilburg University Press.

Bergsma, W., Croon, M., and Hagenaars, J. A. (2009). *Marginal Models: For Dependent, Clustered, and Longitudinal Categorical Data*. New York: Springer.

Greenland, S. (1989). Generalized Mantel-Haneszel estimators for $K\ 2 \times J$ tables. *Biometrics* 45, 183-191.

Lang, J. B., and Agresti, A. (1994). Simultaneously modelling the joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, 89, 625-632.

Lee, A. J. (1993). Generating random binary deviates having fixed marginal distributions and specified degrees of associations. *American Statistician*, 47, 209-215