Thai Journal of Development Administration, Vol.32, No.2 April - June 1992

Number of Connected Design for Two - way Classification Model

Pachijanut Siripanich* and

David Birkes**

Consider an additive two - way classification model with all parameters unknown,

$$y_{ijk} \doteq \mu + \alpha_i + \beta_j + e_{ijk} \tag{1}$$

$i = 1..., a$; $j = 1,... b$; and $k = 1,..., n_{ij}$.

It is allowed that $n_{st} = 0$ and when it is, it means that there is no observation of the form y_{stk}. The model can also be written in a matrix form as

$$Y = X\beta = X_0\mu + A\alpha + B\theta \tag{2}$$

The design matrix X is said to be connected if it has maximal rank (rank = a + b - 1). This implies that all α - contrasts and all β - contrasts are estimable and analysis can be done in an "ordinary way".

* Assistance Professor, School of Applied Statistics, NIDA.

** Professor, Department of Statistics, Oregon State University, USA.

Birkes, Dodge and Seely (1976) introduced the R-process as tool for determining if the design is connected by using, for instance, an $a \times b$ incidence matrix $N = (n_{ij})$. According to R-process there is no difference between $n_{ij} = 1$ and $n_{ij} > 1$, so we will focus on binary design in which $n_{ij} = 0$ or 1, for all i, j. In order to obtain a connected design, it is abvious that there must be at least $a + b - 1$ one's in an incidence matrix. Austin (1960) and Scoins (1962), using theory of graphs, proved that

$$c(a,\, b,\, a+b-1) = a^{b-1}\, b^{a-1} \tag{3}$$

where c(a, b, n) denotes number of connected designs with a treatments, b blocks and n occupied cells (cells with 1's). Applying mathematical induction, equation (3) was proved again by Birkes and Dodge (1986).

We now consider in more general case for number of connected design c(a, b, n) where $a + b - 1 \le n \le ab$. First two theorems are the case where a = 2 and 3. Viewing the problem as connected graph makes thing a lot more simple, so recursive formula and then general formula of c(a, b, n) are obtained as in theorem 3 and 4.

Theorem 1 $c(2, b, n) = \binom{b}{n-b} 2^{2b-n}$ for $b+1 \le n \le 2b$ and zero otherwise.

Proof Let x = number of columns of matrix N with 1 occupied cell.

and y = number of columns of matrix N with 2 occupied cells.

So $x + 2y = n$ and $x + y = b$

Solving for x and y we have

$x = 2b - n$

$y = n - b$

Therefore the number of a × b matrices with n - b columns of $\begin{pmatrix}1\\1\end{pmatrix}$ and 2b - n columns of $\begin{pmatrix}1\\0\end{pmatrix}$ or $\begin{pmatrix}0\\1\end{pmatrix}$ are

$$\binom{b}{n-b} 2^{2b-n}$$

Theorem 2 $$c(3,b,n) = \sum_{t=2}^{3} (-1)^{t+1} \underset{\sum k_i = b, \sum i\, k_i = n}{\sum_{k_1} \cdots \sum_{k_t}} \left[\binom{b}{k_1,\ldots,k_t} \prod_{i=1}^{t} \binom{3}{i}^{k_i} \right]$$

$$= \sum_{k_3} \binom{b}{k_3, n-b-2k_3, 2b-n+k_3} 3^{b-k_3} - \binom{b}{n-b} 3^{2b-n+1}$$

where $b \geq 3$, $b + 2 \leq n \leq 3b$ and the summation over k_3 goes from $|\min(0, 2b-n)|$ to $\left[\frac{n-b}{2}\right]$, the greatest integer less than or equal $\frac{n-b}{2}$

Proof Let k_i be number of columns of N with i 1's and (3 - i) 0's where $i = 1, 2, 3$. So

$$\sum_{i=1}^{3} k_i = b \qquad \text{and}$$

$$\sum_{i=1}^{3} i\, k_i = n$$

Solve these two equations and write k_1 and k_2 in terms of k_3 as

$$k_1 = 2b - n + k_3$$

$$k_2 = n - b - 2k_3$$

Since all k_i's are non - negative integers,

$$n - 2b \leq k_3 \leq \frac{n-b}{2}$$

or equivalently,

$$|\min\{0, 2b-n\}| \le k_3 \le \left[\frac{n-b}{2}\right]$$

By means of R - process, an a × b design N with at least one column consists of all 1's is connected provided that there is no column consists of all 0's. So number of connected designs with this property is

$$\sum_{k_3 \ge 1}^{\left[\frac{n-b}{2}\right]} \binom{b}{k_3, n-b-2k_3, 2b-n+k_3} \binom{3}{3}^{k_3} \binom{3}{2}^{n-b-2k_3} \binom{3}{1}^{2b-n+k_3} \tag{4}$$

The first term in the summation is number of ways of constructing a × b matrix with k_3 columns of all 1's, n - b - 2k_3 columns of two 1"s and one 0's, and 2b - n - k_3 columns of one 1's and two 0's. In each columns, number of ways of assigning i 1's and (3 - i) 0's is $\binom{3}{i}$ and k_i of them are done independently, so the rest in the summation are acheived.

What is left out is number of connected a × b designs with no columns of all 3 occupied cells. In other words, consider design matrix with x columns of one 1's and two 0's and y columns of two 1's and one 0's. We have seen in the proof of Theorem 1 that x and y are

$$x = 2b - n$$
$$y = n - b$$

Number of all possible ways of constructing such matrix is

$$\binom{b}{n-b}\binom{3}{2}^{n-b}\binom{3}{1}^{2b-n}$$

which can be written as

$$\binom{b}{k_3, n-b-k_3, 2b-n+2k_3}\binom{3}{3}^{k_3}\binom{3}{2}^{n-b-k_3}\binom{3}{1}^{2b-n+2k_3} \tag{5}$$

where $k_3 = 0$

Unconnected design matrix occurs when those 0's of "y - column" are all on the same row. There are

$$\binom{b}{n-b}\binom{3}{2}\binom{3}{1}^{2b-n} = \binom{b}{n-b}3^{2b-n+1} \tag{6}$$

such unconnected design matrices. So (4) + (5) - (6) gives desired result.

Denote N(a, b, n) as a total number of a × b matrices with n ones and the rest are zeros, $n \leq ab$. Similarly to (4), it is easy to see that

$$N(a, b, n) = \underset{\sum_{i=1}^{a} k_i = b,\ \sum_{i=1}^{a} i\, k_i = n}{\sum_{k_1}\cdots\sum_{k_a}}\binom{b}{k_1,\ldots k_a}\prod_{i=1}^{a}\binom{a}{i}^{k_i} \tag{7}$$

and $$N(a, b, n) = c(a, b, n) + d(a, b, n) \tag{8}$$

where d(a, b, n) denotes a number of unconnected designs.

Turning an $a \times b$ design matrix with n occupied cells to a bipartite graph with a vertices of one color and b vertices of the other, and with n edges can simplify the problem. For example, equation (7) can be written as

$$N(a,b,n) = \binom{ab}{n}$$

Fortunately, connected design and connected graph are corresponding. The following recursive formula is found.

Theorem 3

$$c(a,b,n) = \binom{ab}{n} - \sum_{\substack{i=1 \\ (i,j) \neq (a,b)}}^{a} \sum_{j=0}^{b} \sum_{k=0}^{n} \binom{a-1}{i-1} \binom{b}{j} \binom{(a-1)(b-j)}{n-k} c(i, j, k)$$

Proof We only need to show that the second term is d(a, b, n). To be disconnected, at least 1 vertex of a must not be connected with some others vertex.

Choose $i - 1$ ($i = 1,\dots, a$) vertices from $a - 1$ vertices and choose j vertices from b vertices then construct a connected k edges bipartite graphs from these $i - 1$, j vertices. Number of possible graphs are $\binom{a-1}{i-1}\binom{b}{j} c(i,j,k)$ Finally, number of graphs made from the rest $(a - i)$, $(b - j)$ vertices and $n - k$ edges is

$$\binom{(a-i)(b-j)}{n-k}$$

Hence the number of unconnected bipartite graphs is

$$d(a,b,n) = \sum_{\substack{i=1 \\ (i,j) \neq (a,b)}}^{a} \sum_{j=0}^{b} \sum_{k=0}^{n} \binom{a-1}{i-1}\binom{b}{j} c(i,\, j,\, k)\binom{(a-i)(b-j)}{n-k}$$

Note that c(a, b, n) = c(b, a, n). Recursive formula can be written in a symmetric form as follows :

$$c(a,b,n) = \binom{ab}{n} - \sum_{\substack{i=0 \\ (i,j) \neq (a,b)}}^{a} \sum_{j=0}^{b} \sum_{k=0}^{n} \frac{i}{a}\binom{a}{i}\binom{b}{j}\binom{(a-i)(b-j)}{n-k} c(i,\, j,\, k)$$

$$c(a,b,n) = \frac{a}{a+b} c(a,b,n) + \frac{b}{a+b} c(a,b,n)$$

$$= \binom{ab}{n} - \sum_{i=0}^{a} \sum_{j=0}^{b} \sum_{k=0}^{n} \frac{i+j}{a+b}\binom{a}{i}\binom{b}{j}\binom{(a-i)(b-j)}{n-k} c(i,\, j,\, k) \tag{10}$$

Riddell and Uhlenbick (1953) derived the explicit expression of d(p, ℓ), number of unconnected of graphs of p points and ℓ lines. In the same manner modification can be done to get c(a, b, n) of bipartite graph. The result is the following.

<u>**Theorem 4**</u>

$$c(a,b,n) = \sum_{t=1}^{\max\{a,b\}} \frac{(-1)^{t+1}}{t} \sum_{\substack{a_1 \dots a_t,\ b_1 \dots b_t \\ \sum a_j = a,\ \sum b_j = b \\ (a_j,\, b_j) \neq (0,0)}} \cdots \sum \sum \cdots \sum \frac{a!b!}{a_1! \dots a_t!\, b_1! \dots b_t!} \binom{\sum_{i=1}^{t} a_i b_i}{n}$$

<u>**Proof**</u> Let the unconnected configulation of a points of one color, say green, b points of other color, say white, and n lines consist of connected suhgraphs with i points from green and j points from white. Let s be an index which distinguishes between the n_{ij} difference connected graphs of i - j points, and let k_{ijs} be the number of lines in the $(i, j)^{th}$ graph.

Clearly we must have

$$\sum_{i=1}^{a}\sum_{j=1}^{b} i\, n_{ij} = a$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b} j\, n_{ij} = b$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{n_{ij}} k_{ijs} = n$$

To partition a points of green and b points of white into n_{01} of isolated points of white, n_{10} isolated points of green, n_{11} connected pairs of green and white points, n_{12} connected bipartite graphs of one point of green and two points of white, etc., we have.

$$a!\, b! \prod_{i}\prod_{j}\left[\frac{1}{(i!\, j!)^{n_{ij}}}\frac{1}{n_{ij!}}\right] \tag{11}$$

Two notes are made here. First, lower and upper limit of i, j under production are $1 \leq i + j \leq a + b - 1$. Want be emphasize that the upper limit is one less than $a + b$ since we must have at least one point seperated from the rest of the graph. Secondly, the term is divided by $n_{ij}!$ to avoid repeated counting of i - j bipartite graphs. Making each i - j bipartite graphs connected we multiply (11) by

$$\pi_{i}\,\pi_{j}\,\pi_{s=1}^{n_{ij}}\; c(i, j, k_{ijs}) \tag{12}$$

Limits of i, j under productions are as mention above. We then have

$$d(a,b,n) = \sum_{n_{ij}} \cdots \sum \sum_{k_{ijs}} \cdots \sum \left[\left[a!b! \prod_i \prod_j \frac{1}{(i!j!)^{n_{ij}} n_{ij}!} \right] \left[\prod_i \prod_j \prod_{s=1}^{n_{ij}} c(i,j,k_{ijs}) \right] \right]$$

$$1 \le i+j \le a+b-1 \qquad 1 \le i+j \le a+b-1$$

$$\sum_{i=1}^{a} \sum_{j=1}^{b} i n_{ij} = a \quad \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{s=1}^{n_{ij}} k_{ijs} = n$$

$$\sum_{i=1}^{a} \sum_{j=1}^{b} j n_{ij} = b \tag{13}$$

We can remove these restictions by multiplying (13) by $x^a y^b z^n$ and get a generating function

$$F_{a,b}(x,y,z) = \sum_{n_{ij}} \cdots \sum \sum_{k_{ijs}} \cdots \sum \left[\left[\prod_i \prod_j \frac{1}{(i!j!)^{n_{ij}} n_{ij}!} \right] \left[\prod_i \prod_j \prod_{s=1}^{n_{ij}} c(i,j,k_{ijs})\, x^i y^j z^{k_{ijs}} \right] \right]$$

$$1 \le i+j \le a+b-1 \qquad 1 \le i+j \le a+b-1$$

$$\frac{d(a,b,n)}{a!b!} = \text{coefficient of } x^a y^b z^n \text{ in } F_{a,b}(x,y,z)$$

Let N(x, y, z) be a generating function of all possible bipartite graph. It is oviously defined by

$$N(x,y,z) = \sum_{n_{ij}} \cdots \sum \sum_{k_{ijs}} \cdots \left[\left[\prod_{i=0}^{\infty} \prod_{j=0}^{\infty} \frac{1}{(i!j!)^{n_{ij}} n_{ij}!} \right] \left[\prod_{i=0}^{\infty} \prod_{j=0}^{\infty} \prod_{s=1}^{n_{ij}} c(i,j,k_{ijs})\, x^i y^j z^{k_{ijs}} \right] \right]$$

$$n_{ij} \text{ not all zero} \qquad (i,j) \ne (0,0) \qquad (i,j) \ne (0,0)$$

and $\dfrac{\binom{ab}{n}}{a!b!} = \text{coefficient of } x^a y^b z^n \text{ in } N(x,y,z) \text{ when } (a,b) \ne (0,0)$

Hence we can write N(x, y, z) as :

$$N(x,y,z)=\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{n=0}^{\infty}\binom{ab}{n}\frac{x^a}{a!}\frac{y^b}{b!}z^n \tag{14}$$

$$N(x,y,z)=\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{n=0}^{ab}\frac{x^a}{a!}\frac{y^b}{b!}z^n N(a,b,n)$$

when $(a,b)\neq(0,0)$

$$N(x,y,z)=\sum_{n_{ij}\text{ not all zero}}\cdots\sum\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\left[\frac{1}{(i!\,j!)^{n_{ij}}\,n_{ij}!}\prod_{s=1}^{n_{ij}}\sum_{k_{ijs}=0}^{\infty}\left[c(i,j,k_{ijs})x^i y^j z^{k_{ijs}}\right]\right]$$

$$=\sum_{n_{ij}\text{ not all zero}}\cdots\sum\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\left[\frac{1}{n_{ij}!}\prod_{s=1}^{n_{ij}}\sum_{k=0}^{\infty}\left[c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right]\right]$$

$$=\sum_{n_{ij}\text{ not all zero}}\cdots\sum\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\left[\frac{1}{n_{ij}!}\left[\sum_{k=0}^{\infty}c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right]^{n_{ij}}\right]$$

$$=1-\sum_{n_{ij}}\cdots\sum\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\left[\frac{1}{n_{ij}!}\left[\sum_{k=0}^{\infty}c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right]^{n_{ij}}\right]$$

$$1+N(x,y,z)=\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\sum_{n_{ij}=0}^{\infty}\left[\frac{1}{n_{ij}!}\left[\sum_{k=0}^{\infty}c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right]^{n_{ij}}\right]$$

$$=\prod_{i=0}^{\infty}\prod_{j=0}^{\infty}\exp\left[\sum_{k=0}^{\infty}c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right]$$

$$=\exp\left[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}c(i,j,k)\frac{x^i}{i!}\frac{y^j}{j!}z^k\right] \tag{15}$$

Let $C(x,y,z)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}c(i,j,k)\frac{x^a}{a!}\frac{y^b}{b!}z^k$ (16)

Obviously $\frac{c(a,b,n)}{a!b!}$ = coefficient of $x^a y^b z^n$ in $c(x,y,z)$ (17)

From (15) and (16) we have

$$
\begin{aligned}
C(x,y,z) &= \log\left[1+N(x,y,z)\right] \\
&= \sum_{t=1}^{\infty} \frac{(-1)^{t+1}}{t}\left[N(x,y,z)\right]^t \\
&= \sum_{t=1}^{\infty} \frac{(-1)^{t+1}}{t}\left[\sum_{\substack{a=0\\(a,b)\neq(0,0)}}^{\infty}\sum_{b=0}^{\infty}\sum_{n=0}^{\infty}\binom{ab}{n}\frac{x^a}{a!}\frac{y^b}{b!}z^n\right]^t \quad \text{, from (14)} \\
&= \sum_{t=1}^{\infty} \frac{(-1)^{t+1}}{t}\left[\sum_{a_1}\cdots\sum_{a_t}\sum_{b_1}\cdots\sum_{b_t}\sum_{n_1}\cdots\sum_{n_t}\binom{a_1b_1}{n_1}\cdots\binom{a_tb_t}{n_t}\frac{x^{\Sigma a_i}}{a_1!\ldots a_t!}\frac{y^{\Sigma b_i}}{b_1!\ldots b_t!}z^{\Sigma n_i}\right] \\
&\qquad (a_i,b_i)\neq(0,0)
\end{aligned}
$$

From (17) we have

$$
\frac{C(x,y,z)}{a!b!} = \sum_{t=1}^{\infty}\frac{(-1)^{t+1}}{t}\sum_{\substack{a_1\\ \sum a_i=a\\ (a_ib_i)\neq(0,0)}}\cdots\sum_{a_t}\sum_{\substack{b_1\\ \sum b_i=b}}\cdots\sum_{b_t}\left[\left[\sum_{\substack{n_1\\ \sum n_i=n}}\cdots\sum_{n_t}\prod_{i=1}^{t}\binom{a_ib_i}{n_i}\right]\prod_{i=1}^{t}\frac{1}{a_i!b_i!}\right]
$$

The rest is to show that

$$
\sum_{n_1}\underset{\sum n_i=n}{\cdots}\sum_{n_t}\prod_{i=1}^{t}\binom{a_ib_i}{n_i} = \binom{\sum_{i=1}^{t}a_ib_i}{n}
$$

This follows the results from Riddell & Uhlenbeck (1953).

Reference

D. Birkes, Dodge and J. Seely (1976). Spanning Sets for Estimable contrasts in Classification Models. The Ann. Stats., Vol.4, No.1, 86-107.

T.L. Austin (1960). The Enumeration of Point Labelled Chromatic Graphs and Trees, Can. J. Math, Vol 12, No. 4, 535-545.

H.I. Scoins (1962). The Number of tree with Nodes of Alternate Parity, Cambridge Phil. Soc., Proc, Vol. 58, 12-16.

T.L Austin (1960), The Enumeration of Point Labelled Chromatic Graphs and Trees, Can. J. of Math. Vol XII # 4 1960 p. 535 - 545.

Nettleton (1960), Some Genalized Theorem on Connectivity, Can. J. of Math Vol. XII # 4 1960. p. 546-554.

D. Birkes and Y. Dodge (1986). The Number of Minimally Connected Block Designs, Computational Statistics and Data Analysis 4 (1986), 269-275, North-Halland.