A Combinatorial Problem Arising in DNA Sequence Alignments
July 7, 1997

Consider the following two sequences of nucleotides

G A A C T G A T and G A C T C A T

where

A : Adenine
T : Thymine
G : Guanine
C : Cytosine.

Is there any evidence that these two sequences have a common evolutionary ancestor? What would you look for to establish such a link?

A DNA sequence can evolve in one of three ways

- insertions
- deletions
- mutations

For example, if we align the given two sequences as:

1	2	3	4	5	6	7	8
G	A	A	C	T	G	A	T
G	Ø	A	C	T	C	A	T
	↓				↓		
	deletion				mutation		

then we say a deletion has occurred in position 2 and a mutation has occurred in position 6.

As a basic principle, a sequence alignment with “few" insertions, deletions, or mutations is evidence that two (or more) sequences have an evolutionary ancestor.

We can translate this alignment of these two sequences into a zero-one matrix by replacing each letter with a 1 and each Ø (or blank) with a 0.

In this way,

1	2	3	4	5	6	7	8
G	A	A	C	T	G	A	T
G	Ø	A	C	T	C	A	T

translates to

1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1

Note that such a 0-1 matrix uniquely identifies a sequence alignment.

Alternatively, we could have aligned the same two sequences as:

1	2	3	4	5	6	7	8	9
G	A	A	C	T	G	Ø	A	T
G	Ø	A	C	T	Ø	C	A	T
	↓				↓	↓		
	deletion				del	insertion		

Associated Zero-One Matrix

1	1	1	1	1	1	0	1	1
1	0	1	1	1	0	1	1	1

Of these two alignments, which provides the most evidence that the two sequences stem from a common ancestor? Is there another alignment that would provide even more evidence?

The answer to both questions would depend on the relative weight (penalty) assigned to deletions versus insertions versus mutations in an alignment. For the purpose of this talk we will just assume there exists some formula for measuring how well two (or more) sequences are aligned.

Reference:

"General Methods of Sequence Comparison", Waterman, *Bulletin of Mathematical Biology*, Vol. 46, No. 4, 1984, pp 473-500.

The second question naturally leads to the idea of writing an algorithm to search over "all possible alignments" for that alignment that maximizes the above formula, whatever that formula may be.

Dynamic programming algorithms and hashing techniques have been developed to accomplish this.

References:

Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons, Sankoff and Kruskal (Editors), 1983, Addison-Wesley, London.

Mathematical Methods for DNA Sequences, Waterman (Editor), 1988, CRC Press.

The natural question to ask before writing any search algorithm is “How many cases (alignments) are there?". Knowing this we can estimate how long such a search will take.

Consider again the two alignments

and

The first alignment consists of a $\binom{1}{1}$ subsequence of length 1 followed by a $\binom{1}{1}$ subsequence of length 6.

The second alignment consists of a $\binom{1}{1}$ subsequence of length 1 followed by a $\binom{1}{1}$ subsequence of length 3 followed by a $\binom{1}{1}$ subsequence of length 2.

Biologists generally find an alignment more believable when the $\binom{1}{1}$ subsequences occur in longer blocks.

Therefore, we now add the restriction that $\binom{1}{1}$ subsequences must occur in blocks of length r or more. If, for example, $r = 2$, then neither of the two alignments should be considered. But we would still consider the alignments

1	2	3	4	5	6	7	8	9
Ø	G	A	A	C	T	G	A	T
G	Ø	Ø	A	C	T	C	A	T
								6

Associated Zero-One Matrix

0	1	1	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1
								6

and

1	2	3	4	5	6	7	8	9	10
G	Ø	A	A	C	T	G	Ø	A	T
Ø	G	Ø	A	C	T	Ø	C	A	T
				3				2	

Associated Zero-One Matrix

1	0	1	1	1	1	1	0	1	1
0	1	0	1	1	1	0	1	1	1
				3				2	

Nothing in what we have done limits us to looking at only two sequences at a time. For example, we can consider the three sequences G A A C T G A T, G A C T C A T, and A G A T.

Ø	G	A	A	C	T	G	A	T
G	Ø	Ø	A	C	T	C	A	T
A	Ø	Ø	Ø	Ø	Ø	G	A	T
						3		

0	1	1	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1
1	0	0	0	0	0	1	1	1
						3		

In this alignment there is only one subsequence of $\begin{bmatrix}1\\1\\1\end{bmatrix}$ (i.e. a column of all "1's") and it has length 3.

As we mentioned before the associated 0-1 matrix uniquely identifies a sequence alignment.

This remains true when aligning more than 2 sequences as well.

What restrictions are placed on this 0-1 matrix from the nature of its association with DNA sequence alignments?

(1) There cannot be any columns consisting of all 0's.

(2) Columns consisting of all 1's must occur in blocks of length $\geq r$.

(3) Each row sum must equal the number of nucleotides in that row.

(4) The number of columns must be greater than or equal to the maximum number of nucleotides in any row.

(5) The number of columns must be less than or equal to the total number of nucleotides in all rows.

The purpose of this paper is to answer the question, "**How many 0-1 matrices are there subject to all of these five restrictions?**"

Let

$t_j =$ number of nucleotides in the j^{th} row

$m =$ the number of sequences being compared (i.e. the number of rows)

$r =$ the minimum acceptable length of contiguous columns of all 1's.

Using this notation we can state that $v = \max(t_1, t_2, \dots, t_m)$ and $T = t_1 + t_2 + \cdots + t_m$.

What results are already known about the number of such 0-1 matrices?

$m = 2$ and $r = 1$, exact solution

$$\sum_{n=0}^{T-v} \binom{t_1}{n} \binom{t_2}{n} 2^n$$

"Asymptotic Limits for a Two-Dimensional Recursion", H. Turner Laquer, *Studies in Applied Mathematics*, 64 : 271-277 , 1981.

$m = 2$ and $t_1 = t_2 = t$, asymptotic approximation as $t \to \infty$

If we let $g(t, r)$ represent the exact count in this case, then

$$g(t,r) \sim \frac{\rho^r - \rho + 1}{\rho^t \sqrt{-\pi \cdot \rho \cdot t \cdot h'(\rho)}} \quad \text{as } t \to \infty$$

where ρ is the smallest positive real root of $h(x) = 0$ where

$$h(x) = (1-x)^2 - 4x(x^r - x + 1)^2.$$

"Sequence Alignments with Matched Sections", Griggs, Hanlon, and Waterman, *SIAM Journal of Algebra and Discrete Mathematics*, Vol. 7, No. 4, October 1986, pp. 604-608.

$r = 1$ and $t_1 = \cdots = t_m = t$, asymptotic approximation as $t \to \infty$

If we let $f(t, m)$ represent the exact count in this case, then

$$f(t,m) = \left(\frac{s^t}{t^{(m-1)/2}}\right)\left(\frac{2^{(m^2-1)/2m}}{\rho\pi^{(m-1)/2}\sqrt{m}} + O\left(\frac{1}{\sqrt{t}}\right)\right)$$

where $\rho = 2^{1/m} - 1$ and $s = \rho^{-m}$.

"On the Number of Alignments of k Sequences", Griggs, Hanlon, Odlyzko, and Waterman, *Graphs and Combinatorics*, Vol. 6, 1990, pg. 133-146.

Note: In their paper they use k where we are using m to represent the number of sequences being compared.

Main result of this paper.

General Case, Exact Solution

For general

$t_j =$ number of nucleotides in the j^{th} row
$m =$ the number of sequences being compared (i.e. the number of rows)
$r =$ the minimum acceptable length of contiguous columns of all 1's.

there are

$$\sum_{n=\nu}^{T}\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k}(-1)^{t-i-l}\binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}\binom{k}{l}$$

$$\times\binom{l}{t_1-i-k+l}\times\cdots\times\binom{l}{t_1-i-k+l}$$

such zero-one matrices where again

$$T = t_1 + t_2 + \cdots + t_m \quad \text{and} \quad \nu = \max(t_1, t_2, \ldots, t_m).$$

Some simplification is possible in special cases.

$m = 2$, exact solution

$$\sum_{n=\nu}^{T} \sum_{j=0}^{\left\lfloor \frac{t_1+t_2-n}{r} \right\rfloor} \binom{j-1+t_1+t_2-n-jr}{t_1+t_2-n-jr} \binom{2n-t_1-t_2+1}{2n-t_1-t_2+1-j} \binom{2n-t_1-t_2}{n-t_2}$$

$r = 1$, exact solution

$$\sum_{n=\nu}^{T} \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} \binom{i}{t_1} \cdots \binom{i}{t_m}.$$

To prove the above general case formula it will be necessary to apply a separate new result that we call the **multidimensional Bernoulli randomization theorem**. We will state this result here but defer its proof to an appendix.

As a setup for this new result, consider a table of m rows such that the j^{th} row has n_j columns, $j = 1, \dots, m$. Suppose that for each $j = 1, \dots, m$, t_j balls are randomly distributed into the n_j cells of that row subject to the restriction that a cell can hold at most one ball. Assume all distributions of balls in a given row are equally likely and that all balls in the table are identical. Let

$$X_{i,j} = \begin{cases} 1 & \text{cell } (i,j) \text{ contains a ball} \\ 0 & \text{else.} \end{cases}$$

That is, assume

$$P\begin{pmatrix} (X_{1,1}, \dots, X_{1,n_1}) = (x_{1,1}, \dots, x_{1,n_1}) \\ \vdots \\ (X_{m,1}, \dots, X_{m,n_m}) = (x_{m,1}, \dots, x_{m,n_m}) \end{pmatrix} = \frac{1}{\binom{n_1}{t_1}} \cdots \frac{1}{\binom{n_m}{t_m}}$$

for all $\left((x_{1,1}, \dots, x_{1,n_1}), \dots, (x_{m,1}, \dots, x_{m,n_m})\right)$ such that $x_{1,1} + \cdots + x_{1,n_1} = t_1, \dots, x_{m,1} + \cdots + x_{m,n_m} = t_m$ and $x_{i,j} \in \{0,1\}$, for all $i = 1, \dots, m$, $j = 1, \dots, n_m$.

With this setup we can state the following result.

Multidimensional Bernoulli Randomization Theorem (proof in an appendix)

$$\mathsf{E}\left(\Psi\left(\left(X_{1,1},\ldots,X_{1,n_1}\right),\ldots,\left(X_{m,1},\ldots,X_{m,n_m}\right)\right)\right)$$

$$=\frac{(n_1-t_1)!\cdots(n_m-t_m)!}{n_1!\cdots n_m!}\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}[\bigstar]\Big|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

where

$$\bigstar\ =(\theta_1+1)^{n_1}\cdots(\theta_m+1)^{n_m}\,\mathsf{E}\left(\Psi\left(\left(Z_{1,1},\ldots,Z_{1,n_1}\right),\ldots,\left(Z_{m,1},\ldots,Z_{m,n_m}\right)\right)\right)$$

and the $Z_{i,j}$ are independent with $Z_{i,j}\sim$ Bernoulli $\left(\frac{\theta_i}{\theta_i+1}\right)$ for all $i=\{1,\ldots,m\}$ and $j\in\{1,\ldots,n_i\}$. That is,

$$P\left(Z_{i,j}=z\right)=\frac{(\theta_i)^z}{\theta_i+1}\qquad z\in\{0,1\}.$$

Now we are position to proof the stated formula for the number of zero-one matrices that will satisfy all five of the stated restrictions.

Proof

Consider each entry in an $m\times n$ matrix as an urn. Suppose we randomly distribute t_i identical balls into the i^{th} row of this matrix of urns subject to the restriction that at most 1 ball can go into an urn for $i\in\{1,\ldots,m\}$. Assume that all $\binom{n}{t_i}$ possible allocations of balls into the i^{th} row are equally likely to occur.

Let $X_{i,j}\in\{0,1\}$ equal the number of balls in the (i,j) urn.

Let $\mathbb{S}^{m,n}_{t_1,\ldots,t_m}$ be the set of all possible values of the vector

$$\left(\left(X_{1,1},\ldots,X_{1,n}\right),\ldots,\left(X_{m,1},\ldots,X_{m,n}\right)\right)$$

which by definition will all be equally likely. Clearly,

$$N\left(\mathbb{S}^{m,n}_{t_1,\ldots,t_m}\right)=\binom{n}{t_1}\times\cdots\times\binom{n}{t_m}.$$

Define $\mathcal{W}_{t_1,\ldots,t_m}$ to be that subset of $\mathbb{S}^{m,n}_{t_1,\ldots,t_m}$ such that

(i) there are no columns of all 0's, and

(ii) columns of all 1's must occur in contiguous blocks of at least r
(iii) the i^{th} row sum equals t_i, $i = 1, \dots, m$.

Then the problem considered here is to find

$$\sum_{n=\nu}^{T} \binom{n}{t_1} \times \cdots \times \binom{n}{t_m} P\left(\left((X_{1,1}, \dots, X_{1,n}), \dots, (X_{m,1}, \dots, X_{m,n})\right) \in \mathcal{W}_{t_1,\dots,t_m}\right).$$

We can apply the multidimensional Bernoulli randomization theorem to find

$$P\left(\left((X_{1,1}, \dots, X_{1,n}), \dots, (X_{m,1}, \dots, X_{m,n})\right) \in \mathcal{W}_{t_1,\dots,t_m}\right)$$

because we are dealing with zero-one matrices where all possible matrices with given row sums are assumed to be equally likely.

Notice that the in the multidimensional Bernoulli randomization theorem it is possible for each row to have a different number of columns. But for this problem n, the number of columns in our zero-one matrix, is the same for all rows. That is, $n_1 = n_2 = \cdots = n_m = n$.

Let $\mathbb{S}^{m,n}$ be the product space

$$\left(\{0,1\}, \dots, \{0,1\}\right) \times \cdots \times \left(\{0,1\}, \dots, \{0,1\}\right)$$

and define $\mathcal{W}$ to be that subset of $\mathbb{S}^{m,n}$ such that

(i) there are no columns of all 0's, and
(ii) columns of all 1's must occur in contiguous blocks of at least r.

That is,

$$\mathcal{W}_{t_1,\dots,t_m} = \mathcal{W} \cap \mathbb{S}^{m,n}_{t_1,\dots,t_m}.$$

Note: $\mathcal{W}$ has the same properties as $\mathcal{W}_{t_1,\dots,t_m}$ but <u>without</u> the restriction that the row sums have to be $t_1, \dots, t_m$.

Now let

$$\Psi\left((X_{1,1}, \dots, X_{1,n}), \dots, (X_{m,1}, \dots, X_{m,n})\right) = \mathbb{I}\left(\left((X_{1,1}, \dots, X_{1,n}), \dots, (X_{m,1}, \dots, X_{m,n})\right) \in \mathcal{W}_{t_1,\dots,t_m}\right)$$

so that

$$\mathsf{E}\left(\Psi\left(\left(X_{1,1},\ldots,X_{1,n}\right),\ldots,\left(X_{m,1},\ldots,X_{m,\mathrm{n}}\right)\right)\right)$$
$$= P\left(\left(\left(X_{1,1},\ldots,X_{1,n}\right),\ldots,\left(X_{m,1},\ldots,X_{m,n}\right)\right)\in \mathcal{W}_{t_1,\ldots,t_m}\right).$$

Applying the multdimensional Bernoulli randomization theorem, it follows that

$$P\left(\left(\left(X_{1,1},\ldots,X_{1,n}\right),\ldots,\left(X_{m,1},\ldots,X_{m,n}\right)\right)\in \mathcal{W}_{t_1,\ldots,t_m}\right)$$
$$= \mathsf{E}\left(\Psi\left(\left(X_{1,1},\ldots,X_{1,n}\right),\ldots,\left(X_{m,1},\ldots,X_{m,n}\right)\right)\right)$$
$$= \frac{(n-t_1)!\cdots(n-t_m)!}{n!\cdots n!}\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}[\bigstar]\Big|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

where

$$\bigstar = (\theta_1+1)^n\cdots(\theta_m+1)^n\,\mathsf{E}\left(\Psi\left(\left(Z_{1,1},\ldots,Z_{1,n}\right),\ldots,\left(Z_{m,1},\ldots,Z_{m,n}\right)\right)\right)$$
$$= (\theta_1+1)^n\cdots(\theta_m+1)^n\,P\left(\left(\left(Z_{1,1},\ldots,Z_{1,n}\right),\ldots,\left(Z_{m,1},\ldots,Z_{m,n}\right)\right)\in \mathcal{W}\right)$$

with $Z_{i,j}\sim$ Bernoulli $\left(\frac{\theta_i}{\theta_i+1}\right)$ and the $Z_{i,j}$ are independent for all $i=\{1,\ldots,m\}$ and $j\in\{1,\ldots,n\}$. That is,

$$P\left(Z_{i,j}=z\right)=\frac{(\theta_i)^z}{\theta_i+1}\qquad z\in\{0,1\}.$$

Now we can concentrate on finding

$$P\left(\left(\left(Z_{1,1},\ldots,Z_{1,n}\right),\ldots,\left(Z_{m,1},\ldots,Z_{m,n}\right)\right)\in \mathcal{W}\right)$$

where $\mathcal{W}$ is that set of $m\times n$ zero-one matrices such that

(i) there are no columns of all 0's, and
(ii) columns of all 1's occur in contiguous blocks of length at least r.

Notice that for this probability calculation

(i) the row sums are *not* fixed

(ii) $Z_{i,j} = 1$ corresponds to the event that a ball is put into cell (i, j) and $Z_{i,j} = 0$ corresponds to the event that a ball is not put into cell (i, j)

(iii) the n columns are ***independent*** random vectors $\big($because the $Z_{i,j}$ are independent for all $i = \{1, \dots, m\}$ and $j \in \{1, \dots, n\}\big)$

(iv) the n columns are ***identically distributed*** random vectors.

The last line (iv) is important to understand. The *rows* of a matrix in $\mathcal{W}$ are <u>not</u> identically distributed because cells in different rows are Bernoulli random variables *with different parameters*. But the *columns* of a matrix in $\mathcal{W}$ contain one cell from each row and hence the columns <u>are</u> identically distributed.

We will label a column as a type A column if it consists of all 1's, a type B column if it consists of neither all 1's nor all 0's and a type C column if it consists of all 0's.

To determine if a given random matrix $\Big(\big(Z_{1,1}, \dots, Z_{1,n}\big), \dots, \big(Z_{m,1}, \dots, Z_{m,n}\big)\Big)$ belongs to the set of matrices $\mathcal{W}$ it is sufficient to know the type labels (A, B or C) of each column. Furthermore, we can determine the probability that a column belongs to each of these types. For each $j = 1,2, \dots, m$ we have

$$P(\text{column } j \text{ is type } A) = P\Big(\big(Z_{1,j}, Z_{2,j}, \dots, Z_{m,j}\big) = (1,1, \dots,1)\Big)$$
$$= \left(\frac{\theta_1}{\theta_1 + 1}\right)\left(\frac{\theta_2}{\theta_2 + 1}\right) \cdots \left(\frac{\theta_m}{\theta_m + 1}\right)$$

$$P(\text{column } j \text{ is type } C) = P\Big(\big(Z_{1,j}, Z_{2,j}, \dots, Z_{m,j}\big) = (0,0, \dots,0)\Big)$$
$$= \left(\frac{1}{\theta_1 + 1}\right)\left(\frac{1}{\theta_2 + 1}\right) \cdots \left(\frac{1}{\theta_m + 1}\right)$$

$$P(\text{column } j \text{ is type } B) = 1 - P(\text{column } j \text{ is type } A) - P(\text{column } j \text{ is type } C).$$

We can see that

$$P\Big(\Big(\big(Z_{1,1}, \dots, Z_{1,n}\big), \dots, \big(Z_{m,1}, \dots, Z_{m,n}\big)\Big) \in \mathcal{W}\Big)$$
$$= P(\text{no } C\text{'s occur and all } A\text{'s occur in strings of length at least } r).$$

It will be useful to express this in the equivalent form

$$P(\text{no } C\text{'s occur and all } A\text{'s occur in strings of length at least } r)$$
$$= P(\text{all } A\text{'s occur in strings of length at least } r|\text{no } C\text{'s occur})P(\text{no } C\text{'s occur}).$$

To be clear, we are using "an A occurs" for the event that a column is of type A (it consists of all 1's), "a B occurs" for the event that a column is neither all 1's nor all 0's and "a C occurs" for the event that a column is all 0's.

Furthermore it will be useful to partition the event "all A's occur in strings of length at least r" into those (disjoint) regions according to the total number of A's and the exact number of disjoint strings of A's (each of length at least r).

In this way we have

$P(\text{no } C\text{'s occur and all } A\text{'s occur in strings of length at least } r)$

$= P(\text{all } A\text{'s occur in strings of length at least } r|\text{no } C\text{'s occur})P(\text{no } C\text{'s occur})$

$$= \sum_{\substack{\text{all } (i,j) \text{ for} \\ \text{fixed } (n,r)}} \sum P(\text{exactly } i\ A\text{'s consisting of exactly } j \text{ strings each of length} \geq r|\text{no } C\text{'s})P(\text{no } C\text{'s})$$

where $\sum_{\substack{\text{all } (i,j) \text{ for} \\ \text{fixed } (n,r)}} \sum$ is our notation for "sum over all feasible values of i and j for the given values of n and r.

Because there are exactly i A's and no C's occur, there must be $n - i$ B's. To be clear, A and B are the only possible column labels.

Now define p as the probability that any particular column of a random matrix in $\mathcal{W}$ has label A *given that* the label is not C. That is, let $p = P(A|\text{not } C)$.

Then $P(B|\text{not } C) = 1 - P(A|\text{not } C) = 1 - p$.

Because columns are independent and identically distributed random vectors the column labels are independent and identically distributed random variables.

Therefore every *feasible* arrangement of i A's and $n-i$ B's (conditional on no C's occuring) has probability $p^i(1-p)^{n-i}$. Hence

$$P(\text{no } C\text{'s occur and all } A\text{'s occur in strings of length at least } r)$$

$$= \sum_{\substack{\text{all } (i,j) \text{ for} \\ \text{fixed } (n,r)}} \sum N_{AB}(n,i,j,r)\, p^i(1-p)^{n-i} P(\text{no } C\text{'s occur})$$

where $N_{AB}(n,i,j,r)$ is the number of feasible arrangements of i A's and $(n-i)$ B's where the A's consist of exactly j strings each of length at least r.

At this point we have a choice to make. One option is to carefully delineate all necessary cases and their appropriate limits of our double summation "over all feasible values of i and j for the given values of n and r". If we choose this option, then we don't have to worry if a formula for $N_{AB}(n,i,j,r)$ is applicable outside of the feasible region.

The second option is to make sure a formula for $N_{AB}(n,i,j,r)$ equals 0 for all (i,j) outside of the feasible region. If we choose this option, then we don't have to worry if our limits of summation include values of (i,j) outside of the feasible region.

This type of option comes up often in combinatorial problems and while there is no general rule, the second option is often leads to "cleaner looking" answers. It is the option we will take for this problem.

But before we derive a formula for $N_{AB}(n,i,j,r)$ we will solve for $p = P(A|\text{not } C)$ and for $P(\text{no } C\text{'s occur})$. We see that

$$p = P(A|\text{not } C) = \frac{P(A \cap \overline{C})}{P(\overline{C})} = \frac{P(A)}{P(\overline{C})} = \frac{\left(\frac{\theta_1}{\theta_1+1}\right)\left(\frac{\theta_2}{\theta_2+1}\right)\cdots\left(\frac{\theta_m}{\theta_m+1}\right)}{1-\left(\frac{1}{\theta_1+1}\right)\left(\frac{1}{\theta_2+1}\right)\cdots\left(\frac{1}{\theta_m+1}\right)}$$

$$= \frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}$$

and because the column label random variables are independent we have

$$P(\text{no } C\text{'s occur}) = \prod_{k=1}^{n} P(\text{column } k \text{ is not type } C)$$

$$= \prod_{k=1}^{n} \Big(1 - P(\text{column } k \text{ is type } C)\Big)$$

$$= \Big(1 - P(\text{column 1 is type } C)\Big)^n$$

$$= \left(1 - \left(\frac{1}{\theta_1 + 1}\right)\left(\frac{1}{\theta_2 + 1}\right)\cdots\left(\frac{1}{\theta_m + 1}\right)\right)^n$$

$$= \frac{\big((\theta_1 + 1)(\theta_2 + 1)\cdots(\theta_m + 1) - 1\big)^n}{\big((\theta_1 + 1)(\theta_2 + 1)\cdots(\theta_m + 1)\big)^n}.$$

We will now show how to construct the set of all such $N_{AB}(n, i, j, r)$ arrangements of A's and B's using count independent steps (so the rule of product applies).

Step 1. Set out j (empty) urns in a row and label them as A_1 to A_j in that order.
Step 2. Place an (empty) urn before the first "A" urn, between each "A" urn, and after the last "A" urn and label them as B_1 to B_{j+1} in that order.
Step 3. Put r (identical) A's in each of urns A_1 to A_j.
Step 4. Put a single B in each of urns B_2 to B_j (leave urns B_1 and B_{j+1} empty)
Step 5. Distribute the remaining $i - rj$ identical A's into urns A_1 to A_j with no restrictions on the number of remaining A's per urn.
Step 6. Distribute the remaining $(n - i) - (j - 1)$ B's into urns B_1 to B_{j+1} with no restrictions on the number of remaining B's per urn.

When putting together a construction of count independent steps you need to justify to yourself that these steps are truly count independent (*i.e.* the number of ways to complete a given step does not depend on the outcome of any previous step), that completing these steps in all possible ways will generate every element in $N_{AB}(n, i, j, r)$ exactly once and that completing these steps in all possible ways cannot generate an element that does not belong to $N_{AB}(n, i, j, r)$.

Once we mentally satisfy ourselves that our six-step construction meets all of these requirements then we have

$$N_{AB}(n, i, j, r) = \prod_{k=1}^{6} (\text{number of ways to do Step } k).$$

There is only one distinct way to accomplish each of steps 1,2,3 and 4. We notice that Step 5 and Step 6 are *multiset* problems (counting the number of ways to put identical balls into labeled urns with no restrictions on the number of balls per urn).

From elementary combinatorics we know that there are $\binom{x+y-1}{y}$ distinct ways to distribute y identical balls into x labeled (distinct) urns.

Hence there are

$$\binom{j+(i-rj)-1}{i-rj} = \binom{i-rj+j-1}{i-rj}$$

ways to accomplish Step 5 and there are

$$\binom{(j+1)+\big((n-i)-(j-1)\big)-1}{(n-i)-(j-1)} = \binom{n-i+1}{n-i+1-j}$$

ways to accomplish Step 6.

Therefore,

$$N_{AB}(n,i,j,r) = \binom{i-rj+j-1}{i-rj}\binom{n-i+1}{n-i+1-j}.$$

Does this formula really equal 0 for all (i,j) outside of the feasible region? The typical definition in counting applications, and the one we are assuming here, for the binomial coefficient $\binom{a}{b}$ defined *over the integers* (positive and negative) is

$$\binom{a}{b} = \begin{cases} \dfrac{a!}{b!\,(a-b)!} & 0 \le b \le a \\ 0 & \text{else} \end{cases}$$

with $0! = 1$.

We do caution the reader who might implement the results in this paper in a computer algebra system which might not default to this definition for binomial coefficients (but can be redefined to). In particular both WolframAlpha [ ] and Maxima [ ] return -2 instead of the expected 0 to the query binomial$[-2,-3]$.

But assuming we have adopted the above standard definition for binomial coefficients defined over the integers, does the derived formula for $N_{AB}(n,i,j,r)$ equal 0 outside the feasible region? We will examine some representative cases to show how this comes about.

But first, just to help in clarifying all this notation, we begin by considering the following *feasible* case of $(i,j) = (5,2)$ when $n = 12$ and $r = 2$. One such feasible arrangement is demonstrated by

$$BAABBBBAAABB$$

where we have $i = 5$ A's broken into $j = 2$ strings and each string is at least $r = 2$ long separated by a total of $n - i = 12 - 5 = 7$ B's.

The formula

$$N_{AB}(12,5,2,2) = \binom{5 - 2(2) + 2 - 1}{5 - 2(2)}\binom{12 - 5 + 1}{12 - 5 + 1 - 2} = \binom{2}{1}\binom{8}{6} = 56$$

shows $BAABBBBAAABB$ to be one of 56 feasible arrangements.

On the other hand, $i = 5, j = 3, n = 12, r = 2$ is <u>not</u> feasible because $j = 3$ blocks of A's each containing *at least* $r = 2$ A's would require at least $jr = 3 \cdot 2 = 6$ A's.

In general we need $i \geq jr$ to be feasible. If we compute $N_{AB}(12,5,3,2)$ we get

$$N_{AB}(12,5,3,2) = \binom{5 - 6 + 3 - 1}{5 - 6}\binom{12 - 5 + 1}{12 - 5 + 1 - 3} = \binom{1}{-1}\binom{8}{5} = 0$$

as required.

$$\binom{1}{-1}\binom{8}{5}$$

Clearly the factor $\binom{i - rj + j - 1}{i - rj}$ will equal 0 whenever $i < rj$.

We also notice that $i = 10, j = 4, n = 12, r = 2$ is <u>not</u> feasible even though $10 = i \geq jr = 8$ because there would only be $n - i = 12 - 10 = 2$ B's and it requires at least $j - 1 = 3$ B's to separate the required $j = 4$ contiguous blocks of A's.

In general we need $n - i + 1 - j \geq 0$ to be feasible. If we compute $N_{AB}(12,10,4,2)$ we get

$$N_{AB}(12,10,4,2) = \binom{10 - 2(4) + 2 - 1}{10 - 2(4)}\binom{12 - 10 + 1}{12 - 10 + 1 - 4} = \binom{3}{2}\binom{3}{-1} = 0$$

as required.

The factor $\binom{n - i + 1}{n - i + 1 - j}$ will equal 0 whenever $n - i + 1 - j < 0$.

As one last example, the case $i = 10, j = 0, n = 12, r = 2$ and $i = 2, j = 1, n = 12, r = 4$ is not feasible because the $i = 10$ A's cannot be broken into 0 blocks

Does our formula for $N_{AB}(n,i,j,r)$ equal 0 in each of this example? We note that

$$N_{AB}(12,10,0,2) = \binom{10-2(0)+0-1}{10-2(0)}\binom{12-10+1}{12-10+1-0} = \binom{9}{10}\binom{3}{3} = 0$$

as required.

Taking stock of where we are, we have now shown that

$$P\left(\left((Z_{1,1},\dots,Z_{1,n}),\dots,(Z_{m,1},\dots,Z_{m,n})\right) \in \mathcal{W}\right)$$

$$= P(\text{all } A\text{'s occur in strings of length at least } r | \text{no } C\text{'s}) P(\text{no } C\text{'s occur})$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n} N_{AB}(n,i,j,r)\, p^i (1-p)^{n-i} P(\text{no } C\text{'s occur})$$

where

$$N_{AB}(n,i,j,r) = \binom{i-rj+j-1}{i-rj}\binom{n-i+1}{n-i+1-j}$$

$$p = \frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}.$$

and

$$P(\text{no } C\text{'s occur}) = \frac{\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1\right)^n}{\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)\right)^n}.$$

Now we will work to simplify $p^i(1-p)^{n-i}P(\text{no } C\text{'s occur})$. We have

$$p^i(1-p)^{n-i}P(\text{no } C\text{'s occur})$$

$$= \left(\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}\right)^i \left(1-\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}\right)^{n-i}$$

$$\times\left(1-\frac{1}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)}\right)^n$$

$$= \left(\frac{\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}}{1-\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}} \right)^i$$

$$\times \left(1-\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}\right)^n$$

$$\times \left(1-\frac{1}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)}\right)^n$$

$$= \left(\frac{\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}}{\frac{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}} \right)^i$$

$$\times \left(\frac{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}\right)^n$$

$$\times \left(1-\frac{1}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)}\right)^n$$

$$= \left(\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}\right)^i$$

$$\times \left(\frac{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1}\right)^n$$

$$\times \left(1-\frac{1}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)}\right)^n$$

$$= \left(\frac{\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}\right)^i$$

$$\times \left(\frac{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m}{(\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)}\right)^n$$

$$= \frac{\left(\theta_1\theta_2\cdots\theta_m\right)^i\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}}{\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)\right)^n}.$$

Therefore,

$$\bigstar = (\theta_1+1)^n\cdots(\theta_m+1)^n\,P\left(\left((Z_{1,1},\ldots,Z_{1,n}),\ldots,(Z_{m,1},\ldots,Z_{m,n})\right)\in\mathcal{W}\right)$$

$$= (\theta_1+1)^n\cdots(\theta_m+1)^n\sum_{i=0}^{n}\sum_{j=0}^{n}N_{AB}(n,i,j,r)\,p^i(1-p)^{n-i}\cdot P(\text{no } C\text{'s occur})$$

$$= (\theta_1+1)^n\cdots(\theta_m+1)^n\sum_{i=0}^{n}\sum_{j=0}^{n}N_{AB}(n,i,j,r)$$
$$\times\frac{\left(\theta_1\theta_2\cdots\theta_m\right)^i\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}}{\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)\right)^n}$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n}N_{AB}(n,i,j,r)\left(\theta_1\theta_2\cdots\theta_m\right)^i$$
$$\times\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n}\binom{i-rj+j-1}{i-rj}\binom{n-i+1}{n-i+1-j}$$
$$\times\left(\theta_1\theta_2\cdots\theta_m\right)^i\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}.$$

Now we will expand $\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}$ and substitute back into our result for ★.

$$\left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)-1-\theta_1\theta_2\cdots\theta_m\right)^{n-i}$$

$$= \sum_{k=0}^{n-i} \binom{n-i}{k} \left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1) - \theta_1\theta_2\cdots\theta_m\right)^k (-1)^{n-i-k}$$

$$= \sum_{k=0}^{n-i}\sum_{l=0}^{k} (-1)^{n-i-k+k-l} \binom{n-i}{k}\binom{k}{l} \left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)\right)^l (\theta_1\theta_2\cdots\theta_m)^{i+k-l}.$$

Substituting this expansion back in we find

$$\bigstar = (\theta_1+1)^n \cdots (\theta_m+1)^n \, P\left(\left((Z_{1,1},\ldots,Z_{1,n}),\ldots,(Z_{m,1},\ldots,Z_{m,n})\right) \in \mathcal{W}\right)$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n} \binom{i-rj+j-1}{i-rj}\binom{n-i+1}{n-i+1-j}$$

$$\times \left(\theta_1\theta_2\cdots\theta_m\right)^i \left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1) - 1 - \theta_1\theta_2\cdots\theta_m\right)^{n-i}$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k} (-1)^{n-i-l} \binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}\binom{k}{l}$$

$$\times (\theta_1\theta_2\cdots\theta_m)^{i+k-l} \left((\theta_1+1)(\theta_2+1)\cdots(\theta_m+1)\right)^l.$$

Putting this together we have

$$\sum_{n=\nu}^{T} \binom{n}{t_1} \times \cdots \times \binom{n}{t_m} P\left(\left((X_{1,1},\ldots,X_{1,n}),\ldots,(X_{m,1},\ldots,X_{m,n})\right) \in \mathcal{W}_{t_1,\ldots,t_m}\right)$$

$$= \sum_{n=\nu}^{T} \binom{n}{t_1} \times \cdots \times \binom{n}{t_m} \frac{(n-t_1)!\cdots(n-t_m)!}{n!\cdots n!} \frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}} [\bigstar]\Big|_{\substack{\theta_1=0 \\ \vdots \\ \theta_m=0}}$$

$$= \sum_{n=\nu}^{T} \frac{1}{t_1!\cdots t_m!} \frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}} [\bigstar]\Big|_{\substack{\theta_1=0 \\ \vdots \\ \theta_m=0}}$$

$$= \sum_{n=\nu}^{T}\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k} (-1)^{n-i-l} \binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}$$

$$\times \binom{k}{l}\left(\prod_{b=1}^{m}\frac{1}{t_b!}\left(\frac{\partial^{t_b}}{\partial\theta_b^{t_b}}\left((\theta_b+1)^l(\theta_b)^{i+k-l}\right)\right)\Bigg|_{\theta_b=0}\right).$$

As our final step we will expand and evaluate the given partial derivative. We have

$$\left(\frac{\partial^{t_b}}{\partial\theta_b^{t_b}}\left((\theta_b+1)^l(\theta_b)^{i+k-l}\right)\right)\Bigg|_{\theta_b=0}$$

$$=\left(\frac{\partial^{t_b}}{\partial\theta_b^{t_b}}\left(\sum_{u_b}^{l}\binom{l}{u_b}(\theta_b)^{u_b+i+k-l}\right)\right)\Bigg|_{\theta_b=0}$$

$$=\sum_{u_b=0}^{l}\binom{l}{u_b}\left(\frac{\partial^{t_b}}{\partial\theta_b^{t_b}}\left((\theta_b)^{u_b+i+k-l}\right)\right)\Bigg|_{\theta_b=0}$$

$$=(t_b)!\sum_{u_b=0}^{l}\binom{l}{u_b}\mathbb{I}(t_b=u_b+i+k-l)$$

$$=(t_b)!\binom{l}{t_b-i-k+l}.$$

Inputting this result for our partial derivatives we will have finally reached our conclusion. Namely, the number of $m\times n$ zero-one matrices where

(i) m, the number of rows, is given,
(ii) columns of all 1's must occur in contiguous blocks of length at least r,
(iii) there are no columns of all 0's,
(iv) the i^{th} row sum equals t_i, $i=1,\dots,m$, for given $t_1,\dots,t_m$ and
(v) n, the number of columns, can vary between $\nu=\max(t_1,\dots,t_m)$ and $T=t_1+\cdots+t_m$, inclusive

equals

$$\sum_{n=\nu}^{T}\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k}(-1)^{n-i-l}\binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}$$

$$\times \binom{k}{l}\left(\prod_{b=1}^{m}\frac{1}{t_b!}\left(\frac{\partial^{t_b}}{\partial\theta_b^{t_b}}\left((\theta_b+1)^l(\theta_b)^{i+k-l}\right)\right)\Bigg|_{\theta_b=0}\right)$$

$$=\sum_{n=\nu}^{T}\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k}(-1)^{n-i-l}\binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}$$
$$\times \binom{k}{l}\left(\prod_{b=1}^{m}\frac{1}{t_b!}\left((t_b)!\binom{l}{t_b-i-k+l}\right)\right)$$

$$=\sum_{n=\nu}^{T}\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n-i}\sum_{l=0}^{k}(-1)^{n-i-l}\binom{j-1+i-jr}{i-jr}\binom{n-i+1}{n-i+1-j}\binom{n-i}{k}$$
$$\times\binom{k}{l}\binom{l}{t_1-i-k+l}\times\cdots\times\binom{l}{t_m-i-k+l}.$$

Appendix

Multidimensional Fermi-Dirac Randomization Theorem

Consider a table of m rows such that the j^{th} row has n_j columns, $j = 1, \dots, m$. Suppose that for each $j = 1, \dots, m$, t_j balls are randomly distributed into the n_j cells of that row subject to the restriction that a cell can hold at most one ball. Assume all distributions of balls in a given row are equally likely and that all balls in the table are identical. Let

$$X_{i,j} = \begin{cases} 1 & \text{cell } (i,j) \text{ contains a ball} \\ 0 & \text{else.} \end{cases}$$

That is, assume

$$P\begin{pmatrix} (X_{1,1}, \dots, X_{1,n_1}) = (x_{1,1}, \dots, x_{1,n_1}) \\ \vdots \\ (X_{m,1}, \dots, X_{m,n_m}) = (x_{m,1}, \dots, x_{m,n_m}) \end{pmatrix} = \frac{1}{\binom{n_1}{t_1}} \cdots \frac{1}{\binom{n_m}{t_m}}$$

for all $\left((x_{1,1}, \dots, x_{1,n_1}), \dots, (x_{m,1}, \dots, x_{m,n_m})\right)$ such that $x_{1,1} + \cdots + x_{1,n_1} = t_1, \dots, x_{m,1} + \cdots + x_{m,n_m} = t_m$ and $x_{i,j} \in \{0,1\}$, for all $i = 1, \dots, m$, $j = 1, \dots, n_m$.

In this case we have the following theorem.

$$\mathrm{E}\left(\Psi\left((X_{1,1}, \dots, X_{1,n_1}), \dots, (X_{m,1}, \dots, X_{m,n_m})\right)\right)$$

$$= \frac{(n_1 - t_1)! \cdots (n_m - t_m)!}{n_1! \cdots n_m!} \frac{\partial^{t_1 + \cdots + t_m}}{\partial \theta_1^{t_1} \cdots \partial \theta_m^{t_m}} [\star]\Big|_{\substack{\theta_1 = 0 \\ \vdots \\ \theta_m = 0}}$$

where

$$\star = (\theta_1 + 1)^{n_1} \cdots (\theta_m + 1)^{n_m} \mathrm{E}\left(\Psi\left(\left(Z_{1,1}, \dots, Z_{1,n_1}\right), \dots, \left(Z_{m,1}, \dots, Z_{m,n_m}\right)\right)\right)$$

and the $Z_{i,j}$ are independent with $Z_{i,j} \sim$ Bernoulli $\left(\frac{\theta_i}{\theta_i + 1}\right)$ for all $i = \{1, \dots, m\}$ and $j \in \{1, \dots, n_i\}$. That is,

$$P(Z_{i,j} = z) = \frac{(\theta_i)^z}{\theta_i + 1} \qquad z \in \{0,1\}.$$

Proof

Let $Z_{i,j}$ be independent with $Z_{i,j} \sim$ Bernoulli $\left(\frac{\theta_i}{\theta_i+1}\right)$ for all $i = \{1, \dots, m\}$ and $j \in \{1, \dots, n_i\}$. That is,

$$P(Z_{i,j} = z) = \binom{1}{z}\left(\frac{\theta_i}{\theta_i + 1}\right)^z \left(1 - \frac{\theta_i}{\theta_i + 1}\right)^{1-z}$$

$$= \left(\frac{\theta_i}{\theta_i + 1}\right)^z \left(\frac{1}{1 + \theta_i}\right)^{1-z} = \frac{(\theta_i)^z}{\theta_i + 1}$$

where

$$z \in \{0,1\}, \quad i \in \{1, \dots, m\} \text{ and } j \in \{1, \dots, n_i\}.$$

In this case,

$$P\left(\begin{matrix}(Z_{1,1}, \dots, Z_{1,n_1}) = (z_{1,1}, \dots, z_{1,n_1}) \\ \vdots \\ (Z_{m,1}, \dots, Z_{m,n_m}) = (z_{m,1}, \dots, z_{m,n_m})\end{matrix}\middle|\begin{matrix}Z_{1,1} + \cdots + Z_{1,n_1} = t_1 \\ \vdots \\ Z_{m,1} + \cdots + Z_{m,n_m} = t_m\end{matrix}\right)$$

$$= \frac{P\left(\begin{matrix}(Z_{1,1}, \dots, Z_{1,n_1}) = (z_{1,1}, \dots, z_{1,n_1}) \\ \vdots \\ (Z_{m,1}, \dots, Z_{m,n_m}) = (z_{m,1}, \dots, z_{m,n_m})\end{matrix}\right) \cdot \mathbb{I}\begin{pmatrix} z_{1,1} + \cdots + z_{1,n_1} = t_1 \\ \vdots \\ z_{m,1} + \cdots + z_{m,n_m} = t_m \end{pmatrix}}{P\begin{pmatrix} Z_{1,1} + \cdots + Z_{1,n_1} = t_1 \\ \vdots \\ Z_{m,1} + \cdots + Z_{m,n_m} = t_m \end{pmatrix}}$$

$$= \frac{\left(\left(\frac{(\theta_1)^{z_{1,1}}}{\theta_1 + 1}\right) \cdots \left(\frac{(\theta_1)^{z_{1,n_1}}}{\theta_1 + 1}\right)\right) \cdots \left(\left(\frac{(\theta_m)^{z_{m,1}}}{\theta_m + 1}\right) \cdots \left(\frac{(\theta_m)^{z_{m,n_m}}}{\theta_m + 1}\right)\right) \cdot \mathbb{I}\begin{pmatrix} z_{1,1} + \cdots + z_{1,n_1} = t_1 \\ \vdots \\ z_{m,1} + \cdots + z_{m,n_m} = t_m \end{pmatrix}}{\left(\binom{n_1}{t_1}\left(\frac{\theta_1}{\theta_1 + 1}\right)^{t_1}\left(1 - \frac{\theta_1}{\theta_1 + 1}\right)^{n_1 - t_1}\right) \cdots \left(\binom{n_m}{t_m}\left(\frac{\theta_m}{\theta_m + 1}\right)^{t_m}\left(1 - \frac{\theta_m}{\theta_m + 1}\right)^{n_m - t_m}\right)}$$

$$= \frac{\left(\frac{(\theta_1)^{z_{1,1}+\cdots+z_{1,n_1}}}{(\theta_1+1)^{n_1}}\right)\cdots\left(\frac{(\theta_m)^{z_{m,1}+\cdots+z_{m,n_m}}}{(\theta_m+1)^{n_m}}\right)\cdot \mathbb{I}\begin{pmatrix} z_{1,1}+\cdots+z_{1,n_1}=t_1 \\ \vdots \\ z_{m,1}+\cdots+z_{m,n_m}=t_m \end{pmatrix}}{\binom{n_1}{t_1}\cdots\binom{n_m}{t_m}\frac{(\theta_1)^{t_1}\cdots(\theta_m)^{t_m}}{(\theta_1+1)^{n_1}\cdots(\theta_m+1)^{n_m}}}$$

$$= \frac{\left(\frac{(\theta_1)^{t_1}}{(\theta_1+1)^{n_1}}\right)\cdots\left(\frac{(\theta_m)^{t_m}}{(\theta_m+1)^{n_m}}\right)\cdot \mathbb{I}\begin{pmatrix} z_{1,1}+\cdots+z_{1,n_1}=t_1 \\ \vdots \\ z_{m,1}+\cdots+z_{m,n_m}=t_m \end{pmatrix}}{\binom{n_1}{t_1}\cdots\binom{n_m}{t_m}\frac{(\theta_1)^{t_1}\cdots(\theta_m)^{t_m}}{(\theta_1+1)^{n_1}\cdots(\theta_m+1)^{n_m}}}$$

$$= \frac{1}{\binom{n_1}{t_1}\cdots\binom{n_m}{t_m}}\cdot \mathbb{I}\begin{pmatrix} z_{1,1}+\cdots+z_{1,n_1}=t_1 \\ \vdots \\ z_{m,1}+\cdots+z_{m,n_m}=t_m \end{pmatrix}$$

$$= P\begin{pmatrix} (X_{1,1},\ldots,X_{1,n_1})=(z_{1,1},\ldots,z_{1,n_1}) \\ \vdots \\ (X_{m,1},\ldots,X_{m,n_m})=(z_{m,1},\ldots,z_{m,n_m}) \end{pmatrix}.$$

Note that it follows from this identity that

$$\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\ldots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\ldots,Z_{m,n_m})\end{pmatrix}\middle|\begin{matrix}Z_{1,1}+\cdots+Z_{1,n_1}\\ \vdots \\ Z_{m,1}+\cdots+Z_{m,n_m}\end{matrix}\right) = \mathsf{E}\left(\Psi\begin{pmatrix}(X_{1,1},\ldots,X_{1,n_1})\\ \vdots \\ (X_{m,1},\ldots,X_{m,n_m})\end{pmatrix}\right).$$

for any statistic $\Psi(\ \ )$. Now we can use the general rule of iterated expectations, $\mathsf{E}(X) = \mathsf{E}\big(\mathsf{E}(X|Y)\big)$, to find

$$\mathsf{E}\Big(\Psi\big((Z_{1,1},\ldots,Z_{1,n_1}),\ldots,(Z_{m,1},\ldots,Z_{m,n_m})\big)\Big)$$

in terms of

$$\mathsf{E}\Big(\Psi\big((X_{1,1},\ldots,X_{1,n_1}),\ldots,(X_{m,1},\ldots,X_{m,n_m})\big)\Big).$$

We find that

$$\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\dots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\dots,Z_{m,n_m})\end{pmatrix}\right)=\mathsf{E}\left(\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\dots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\dots,Z_{m,n_m})\end{pmatrix}\middle|\begin{matrix}Z_{1,1}+\cdots+Z_{1,n_1}=s_1\\ \vdots \\ Z_{m,1}+\cdots+Z_{m,n_m}=s_m\end{matrix}\right)\right)$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\left(\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\dots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\dots,Z_{m,n_m})\end{pmatrix}\middle|\begin{matrix}Z_{1,1}+\cdots+Z_{1,n_1}=s_1\\ \vdots \\ Z_{m,1}+\cdots+Z_{m,n_m}=s_m\end{matrix}\right)\right.$$

$$\left.\times\ P\begin{pmatrix}Z_{1,1}+\cdots+Z_{1,n_1}=s_1\\ \vdots \\ Z_{m,1}+\cdots+Z_{m,n_m}=s_m\end{pmatrix}\right)$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\dots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\dots,Z_{m,n_m})\end{pmatrix}\middle|\begin{matrix}Z_{1,1}+\cdots+Z_{1,n_1}=s_1\\ \vdots \\ Z_{m,1}+\cdots+Z_{m,n_m}=s_m\end{matrix}\right)\prod_{j=1}^{m}\binom{n_j}{s_j}\frac{(\theta_j)^{s_j}}{(\theta_j+1)^{n_j}}$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}(X_{1,1},\dots,X_{1,n_1})\\ \vdots \\ (X_{m,1},\dots,X_{m,n_m})\end{pmatrix}\right)\prod_{j=1}^{m}\binom{n_j}{s_j}\frac{(\theta_j)^{s_j}}{(\theta_j+1)^{n_j}}.$$

Therefore,

$$\prod_{j=1}^{m}(1+\theta_j)^{n_j}\cdot\mathsf{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\dots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\dots,Z_{m,n_m})\end{pmatrix}\right)$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}(X_{1,1},\dots,X_{1,n_1})\\ \vdots \\ (X_{m,1},\dots,X_{m,n_m})\end{pmatrix}\right)\prod_{j=1}^{m}\binom{n_j}{s_j}(\theta_j)^{s_j}$$

and

$$\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}\left[\prod_{j=1}^{m}\left(1+\theta_j\right)^{n_j}\cdot \mathsf{E}\left(\Psi\begin{pmatrix}\left(Z_{1,1},\ldots,Z_{1,n_1}\right)\\ \vdots \\ \left(Z_{m,1},\ldots,Z_{m,n_m}\right)\end{pmatrix}\right)\right]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

$$=\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}\left[\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}\left(X_{1,1},\ldots,X_{1,n_1}\right)\\ \vdots \\ \left(X_{m,1},\ldots,X_{m,n_m}\right)\end{pmatrix}\right)\prod_{j=1}^{m}\binom{n_j}{s_j}\left(\theta_j\right)^{s_j}\right]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}\left(X_{1,1},\ldots,X_{1,n_1}\right)\\ \vdots \\ \left(X_{m,1},\ldots,X_{m,n_m}\right)\end{pmatrix}\right)\left(\prod_{j=1}^{m}\binom{n_j}{s_j}\right)\left(\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}\left[\prod_{j=1}^{m}\left(\theta_j\right)^{s_j}\right]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}\right)$$

$$=\sum_{s_1=0}^{n_1}\cdots\sum_{s_m=0}^{n_m}\mathsf{E}\left(\Psi\begin{pmatrix}\left(X_{1,1},\ldots,X_{1,n_1}\right)\\ \vdots \\ \left(X_{m,1},\ldots,X_{m,n_m}\right)\end{pmatrix}\right)\left(\prod_{j=1}^{m}\binom{n_j}{s_j}\right)\left(\prod_{j=1}^{m}\left(t_j!\ \mathbb{I}\left(s_j=t_j\right)\right)\right)$$

$$=\mathsf{E}\left(\Psi\begin{pmatrix}\left(X_{1,1},\ldots,X_{1,n_1}\right)\\ \vdots \\ \left(X_{m,1},\ldots,X_{m,n_m}\right)\end{pmatrix}\right)\left(\prod_{j=1}^{m}\binom{n_j}{t_j}\right)\left(\prod_{j=1}^{m}\left(t_j!\right)\right).$$

It follows that

$$\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}\left[\prod_{j=1}^{m}\left(1+\theta_j\right)^{n_j}\cdot \mathsf{E}\left(\Psi\begin{pmatrix}\left(Z_{1,1},\ldots,Z_{1,n_1}\right)\\ \vdots \\ \left(Z_{m,1},\ldots,Z_{m,n_m}\right)\end{pmatrix}\right)\right]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

$$=\mathsf{E}\left(\Psi\begin{pmatrix}\left(X_{1,1},\ldots,X_{1,n_1}\right)\\ \vdots \\ \left(X_{m,1},\ldots,X_{m,n_m}\right)\end{pmatrix}\right)\left(\prod_{j=1}^{m}\binom{n_j}{t_j}\right)\left(\prod_{j=1}^{m}\left(t_j!\right)\right)$$

or

$$\mathrm{E}\left(\Psi\begin{pmatrix}(X_{1,1},\ldots,X_{1,n_1})\\ \vdots \\ (X_{m,1},\ldots,X_{m,n_m})\end{pmatrix}\right)$$

$$=\left(\prod_{j=1}^{m}\frac{(n_j-t_j)!}{(n_j)!}\right)\cdot\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}\left[\prod_{j=1}^{m}(1+\theta_j)^{n_j}\cdot\mathrm{E}\left(\Psi\begin{pmatrix}(Z_{1,1},\ldots,Z_{1,n_1})\\ \vdots \\ (Z_{m,1},\ldots,Z_{m,n_m})\end{pmatrix}\right)\right]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}.$$

That is,

$$\mathrm{E}\left(\Psi\left((X_{1,1},\ldots,X_{1,n_1}),\ldots,(X_{m,1},\ldots,X_{m,n_m})\right)\right)$$

$$=\frac{(n_1-t_1)!\cdots(n_m-t_m)!}{n_1!\cdots n_m!}\frac{\partial^{t_1+\cdots+t_m}}{\partial\theta_1^{t_1}\cdots\partial\theta_m^{t_m}}[\star]\Bigg|_{\substack{\theta_1=0\\ \vdots \\ \theta_m=0}}$$

where

$$\star=(\theta_1+1)^{n_1}\cdots(\theta_m+1)^{n_m}\,\mathrm{E}\left(\Psi\left((Z_{1,1},\ldots,Z_{1,n_1}),\ldots,(Z_{m,1},\ldots,Z_{m,n_m})\right)\right).$$