

Enumerating Row Arrangements of Three Species

MICHAEL AVIDON
Connecticut College
New London, CT 06320

RICK MABRY
PAUL SISSON
Louisiana State University in Shreveport
Shreveport, LA 71115

Introduction A classic problem in beginning Finite or Discrete Mathematics courses reads as follows: *A photographer wants to arrange N men and N women in a line so that men and women alternate. In how many ways can this be done?*

This problem nicely illustrates the use of factorials and has a simple solution, though students often neglect the factor of two in the answer $2(N!)^2$. This omission can be instructive, as it leads naturally to generalizations of the problem: How does the answer change if there are N men and $N - 1$ women? What if men outnumber women by 2 or more? What if a sexist photographer insists that the lineup start with a man?

These variations are all easily dealt with, and illustrate some of the possible subtleties encountered in counting problems. Another variation of the problem is not, however, so easily handled. The setting (or sitting) must change slightly: *A pet photographer must pose C cats, D dogs, and E emus in a line, with no two animals of the same species adjacent. In how many ways can this be done?*

Skirting the issue of whether emus can in fact be considered pets, we quickly discover many interesting features in this variation of the problem. For example, C , D , and E can differ by more than 1 and still admit nonzero solutions. We also discover that the new problem is quite a bit harder than the original. Dating back to 1966 ([6]), it is often called *Smirnov's problem* and it has practical applications in such areas such as queuing, transportation flow, sequential analysis and multivariate order statistics (see, e.g., [1, 5]). The problem has been solved in a variety of ways (some of which we mention at the end of this note) but unlike those methods, what is presented here is entirely elementary and could be discussed in a first course in

discrete mathematics or combinatorics.

Two ways of counting Consider any *arrangement* of a row of C cats, D dogs and E emus, distinguishable only by species. Let $\mathcal{A}(C, D, E)$ be the set of all such arrangements. As an example, consider the member of $\mathcal{A}(4, 2, 3)$ represented by the string *cccdeeeed*. Obviously this arrangement is not of the sort we are trying to count, as some adjacent seats are occupied by the same species. On the other hand, the arrangement *cdcecedec* in $\mathcal{A}(4, 2, 3)$ is the kind we want. Let $\mathcal{L}(C, D, E)$ denote the set of all such *legal arrangements* in $\mathcal{A}(C, D, E)$. We need to find the cardinality of $\mathcal{L}(C, D, E)$, which we denote by $L(C, D, E)$. Now since these critters are pets, they should all be considered distinct! In that case we simply multiply $L(C, D, E)$ by $C!D!E!$ to account for the permuting of the animals within each species. But the crux of the problem is to find a formula for $L(C, D, E)$, and so we will henceforth, unless otherwise stated, neglect the individuality of the creatures within each species.

Now let \mathcal{A} and \mathcal{L} denote, respectively, the set of all arrangements and legal arrangements, of any size. Define a map from \mathcal{A} into \mathcal{L} as follows: For a given arrangement in \mathcal{A} , replace any run of identical species with a single member of that species (e.g., *cccdeeeed* \rightarrow *cdced*). Then a member of $\mathcal{A}(C, D, E)$ will be mapped to some member of $\mathcal{L}(C', D', E')$, with $1 \leq C' \leq C$, $1 \leq D' \leq D$, and $1 \leq E' \leq E$. In particular, an arrangement is legal iff it is mapped to itself.

We now seek the number of members of $\mathcal{A}(C, D, E)$ that will be mapped to a fixed member of $\mathcal{L}(C', D', E')$. This is a familiar type of pigeonhole problem, as we are asking for the number of ways that C cats can be distributed among C' compartments, and similarly for the dogs and emus. Many readers will recognize that the answer is

$$\binom{C-1}{C'-1} \binom{D-1}{D'-1} \binom{E-1}{E'-1}. \quad (1)$$

For completeness we include the standard “stars and bars” proof.

LEMMA 1. *The number of ways of distributing N indistinguishable items among N' nonempty compartments is $\binom{N-1}{N'-1}$.*

Proof. Consider a line of N “stars”. From among the $N-1$ spaces between the stars, there are $\binom{N-1}{N'-1}$ ways to choose positions for $N'-1$ “bars”, resulting in an arrangement such as:

$$*** | ** | \cdots *** | *$$

The $N'-1$ bars delineate N' nonempty “compartments”, so the proof is complete.

Applying this lemma to each species in turn and invoking the multiplication principle produces formula (1).

Since $L(C', D', E')$ stands for the number of legal arrangements of C' cats, D' dogs, and E' emus, it is clear that

$$\binom{C-1}{C'-1} \binom{D-1}{D'-1} \binom{E-1}{E'-1} L(C', D', E')$$

gives us the number of ways that a chain of $C+D+E$ seats can be filled with C' runs of cats, D' runs of dogs, and E' runs of emus. We thus have two different formulas expressing the number of ways of arranging a chain of C , D , and E indistinguishable cats, dogs, and emus:

$$\sum_{C'=1}^C \sum_{D'=1}^D \sum_{E'=1}^E \binom{C-1}{C'-1} \binom{D-1}{D'-1} \binom{E-1}{E'-1} L(C', D', E') = \frac{(C+D+E)!}{C! D! E!} \quad (2)$$

Inverting the formula Our goal is to invert (2) to determine $L(C, D, E)$. Fortunately, the inversion is easily accomplished due to the nature of the much-studied coefficients in front of the function L .

To begin, we define two matrices (of unspecified size, for the moment) $A = (a_{i,j})$ and $B = (b_{i,j})$, where

$$a_{i,j} = \binom{i-1}{j-1} \quad \text{and} \quad b_{i,j} = (-1)^{i-j} a_{i,j}.$$

(As usual, $\binom{0}{0} = 1$ and $\binom{m}{n} = 0$ if $m < n$.) Although A and B are not necessarily square, the following lemma demonstrates that the two matrices are “inverses” of one another, in a sense.

LEMMA 2. *Assume the sizes of A and B (resp. B and A) are such that the matrix product AB (resp. BA) is defined. Then $\sum_j a_{i,j} b_{j,k} = \delta_{i,k}$ (resp. $\sum_j b_{i,j} a_{j,k} = \delta_{i,k}$).*

Proof. Let n equal the number of columns of A (and hence the number of rows of B). We will prove that $\sum_j a_{i,j} b_{j,k} = \delta_{i,k}$. (The proof of the second claim is similar.)

Let $C = AB$, and note that if we let $C = (c_{i,k})$, then $c_{i,k} = 0$ for $i < k$. For $i \geq k$, observe that

$$\begin{aligned} c_{i,k} &= \sum_{j=1}^n a_{i,j} b_{j,k} = \sum_{j=1}^n \binom{i-1}{j-1} \binom{j-1}{k-1} (-1)^{j-k} \\ &= \binom{i-1}{k-1} \sum_{j=1}^n \binom{i-k}{i-j} (-1)^{j-k} = \binom{i-1}{k-1} \sum_{j=k}^i \binom{i-k}{i-j} (-1)^{j-k}, \end{aligned}$$

where we have used the identity

$$\binom{n}{k} \binom{k}{m} = \binom{n}{m} \binom{n-m}{n-k}$$

and the fact that many of the summands are 0. If $i = k$, then the sum in the last line (and hence the entire expression) is 1, while otherwise it follows from the binomial theorem that the sum is $(1 - 1)^{i-k} = 0$. This completes the proof.

Note, then, that if f and g are arbitrary functions defined on the natural numbers, each of the equations $f(i) = \sum_j a_{i,j} g(j)$ and $g(i) = \sum_j b_{i,j} f(j)$ implies the other. A proof of one implication follows, with the other proved in identical fashion. Assuming the former equation holds, we have

$$\sum_j b_{i,j} f(j) = \sum_j b_{i,j} \sum_k a_{j,k} g(k) = \sum_k g(k) \sum_j b_{i,j} a_{j,k} = \sum_k g(k) \delta_{i,k} = g(i).$$

Moreover, we can repeat the process, so that if f and g are two arbitrary functions defined on triples of natural numbers, we have

$$f(i, j, k) = \sum_{i', j', k'} a_{i, i'} a_{j, j'} a_{k, k'} g(i', j', k') \iff g(i, j, k) = \sum_{i', j', k'} b_{i, i'} b_{j, j'} b_{k, k'} f(i', j', k'). \quad (3)$$

We are now ready to isolate the function L from equation (2).

THEOREM.

$$L(C, D, E) = \sum_{C'=1}^C \sum_{D'=1}^D \sum_{E'=1}^E (-1)^{C+D+E-C'-D'-E'} \times \frac{\binom{C-1}{C'-1} \binom{D-1}{D'-1} \binom{E-1}{E'-1} (C'+D'+E')!}{C'! D'! E'!}.$$

Proof. The formula follows immediately upon applying the observation in equation (3) to equation (2). Note that we are, in effect, “peeling off” the binary coefficients one at a time from the left-hand side of (2), with the result that they appear on the right-hand side along with the appropriate power of -1 .

This finally reveals the photographer’s conundrum. Once we acknowledge the individuality of the pets, the number of different seatings of the C cats, D dogs, and E emus is $L(C, D, E)C! D! E!$.

Just how bad is the situation for the photographer? We list below some values of special interest, namely, when $C = D = E$. Asymptotic estimates for L are known [1, 5]. In particular, $L(N, N, N)$ is asymptotic to a constant multiple of $\frac{8^N}{N}$.

N	1	2	3	4	5	6	7	8	9
$L(N, N, N)$	6	30	174	1092	7188	48852	339720	2403588	17236524

Some sophisticated methods Having presented an elementary approach to the problem, we offer a few tantalizing glimpses at some methods known to specialists. (Our problem can be viewed in many other frameworks. For example, $L(C, D, E)$ is the number of different ways a labeled chain of $C + D + E$ vertices can be properly colored (in the usual graph-theoretic sense) with three colors 1, 2, and 3, each used C , D , and E times, respectively. Or it can be viewed as the number of *words* that can be formed from the *alphabet* $\{c, d, e\}$, without adjacencies, and with the letter c being used C times, etc.)

1. E. Rodney Canfield alerted the authors to a technique called the “transfer matrix method” (see, e.g., [3] or [7]), the result of which is that $L(C, D, E)$ is the coefficient of $x^C y^D z^E$ in the matrix product

$$[x \ y \ z] \begin{bmatrix} 0 & y & z \\ x & 0 & z \\ x & y & 0 \end{bmatrix}^{C+D+E-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

2. Ira Gessel informed the authors of another approach [4, p. 69]: $L(C, D, E)$ is the coefficient of $x^C y^D z^E$ in the power series expansion of

$$\left(1 - \frac{x}{x+1} - \frac{y}{y+1} - \frac{z}{z+1}\right)^{-1}.$$

3. Ira Gessel also related the problem to *rook polynomials* [2, pp. 160–162]. To use this approach, begin by defining

$$L_n(x) = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} \binom{n-1}{k} k! x^{n-k}$$

and let Φ be the linear functional (defined on the usual vector space of polynomials) that sends x^m to $m!$. Then

$$\Phi(L_C(x)L_D(x)L_E(x)) = L(C, D, E) C! D! E!.$$

It is also instructive to apply symbolic software, such as *Mathematica*, to explore all four methods of calculating $L(C, D, E)$. Be forewarned, however, that the first two methods above will require significant computing time for all but small values of C , D and E .

Generalizations The problem can, of course, be generalized. In one direction, each of the formulas presented for $L(C, D, E)$ can easily be extended to handle more species. For example, the number of legal arrangements of C cats, D dogs, E emus, and F frogs is

$$L(C, D, E, F) = \sum_{C'=1}^C \sum_{D'=1}^D \sum_{E'=1}^E \sum_{F'=1}^F (-1)^{C+D+E+F-C'-D'-E'-F'} \times \\ \binom{C-1}{C'-1} \binom{D-1}{D'-1} \binom{E-1}{E'-1} \binom{F-1}{F'-1} \frac{(C'+D'+E'+F')!}{C'! D'! E'! F'!}.$$

The formula can also be restricted to the original simple problem. If only cats and dogs are present, the obvious changes lead to a double summation formula for $L(C, D)$. Of course, we already knew $L(C, D)$: it must be either two, one, or zero, depending on whether $C = D$, $|C - D| = 1$, or $|C - D| \geq 2$, respectively. It is an interesting exercise to verify that the formula in the case of two species really does simplify in this manner.

Further generalizations are left to the reader to explore. For example, one of our students asked for the number of arrangements such that no member of a species is trapped between two other members of the same species. Or perhaps the photographer will permit blocks of up to k members of a given species, but no more. Other variations might concern circular arrangements or two-dimensional grid arrangements of seats. (The latter problem belongs to a class of more general problems concerning the enumeration of n -colorings of labeled graphs with specified numbers of each color, already very difficult for $n = 2$.)

Acknowledgment. The authors thank E. Rodney Canfield, Ira Gessel, Carl Pomerance, and an anonymous reviewer for their help in relating the problem to standard enumeration (emu-neration?) techniques, and for providing references to those methods.

References

- [1] R. Alter and B.P. Lientz, Applications of a generalized combinatorial problem of Smirnov, *Naval Res. Logist.* 16 (1969), 543–546.
- [2] I.M. Gessel, Generalized Rook Polynomials and Orthogonal Polynomials, in *q-Series and Partitions*, IMA Volumes in Mathematics and its Applications 18, Springer-Verlag, New York, NY, 1989, 159–176.

- [3] I.M. Gessel and R.P. Stanley, Algebraic enumeration, in *Handbook of Combinatorics*, ed. R.L. Graham, M. Grottschel, and L. Lovasz, Elsevier (North-Holland), Amsterdam, Holland, 1995, pp. 1021–1061.
- [4] I.P. Goulden and D.M. Jackson, *Combinatorial Enumeration*, Wiley-Interscience, New York, NY, 1983.
- [5] O.V Sarmanov and V.K. Zaharov, A combinatorial problem of Smirnov, *Soviet Math. Dokl.* 9 (1967), 1147–1150.
- [6] N.V. Smirnov, O.V Sarmanov, and V.K. Zaharov, A combinatorial problem of Smirnov, *Soviet Math. Dokl.* 7 (1966), 563.
- [7] R.P. Stanley, *Enumerative Combinatorics*, Vol. 1, 2nd ed. (Cambridge Studies in Advanced Mathematics 49), Cambridge University Press, Cambridge, UK, 1997.