The number of preference orderings: a recursive approach

Ben Eggleston

The Mathematical Gazette / Volume 99 / Issue 544 / March 2015, pp 21 - 32
DOI: 10.1017/mag.2014.3, Published online: 13 March 2015

The number of preference orderings: a recursive approach

BEN EGGLESTON

1. *Introduction*

In modern society, voting is ubiquitous: it determines the outcomes of elections, the rankings of sports teams in many leagues, and the winners of the Oscars. Vast sums of money, not to mention questions of war and peace, turn on the voting systems used in these and other contests. Thus, the formulation of voting systems is a prime example of a mathematical problem that also has substantial real-world implications.

A voting system is, essentially, a function. A simple kind of voting system takes voters' top choices (for example, each voter's favourite candidate for some office) as its input and produces a social choice (the winner of the election) as its output. Here, though, we will understand voting systems more robustly: we will say that a voting system asks each voter not just to indicate his or her top choice, but to rank all of the candidates, from best to worst, and then produces a ranking of all of the candidates as its output. A ranking of all of the alternatives available in some context is called a *preference ordering*. With this term in hand, then, we can say that a voting system is a function that takes a set of individual preference orderings (the voters' respective preference orderings) as its input and produces a preference ordering (hopefully one that deserves to be called the *collective* preference ordering) as its output.

One of the challenges facing the designer of any voting system is to ensure that it will yield an intuitively acceptable output for any possible input; that is, for any possible set of individual preference orderings. To meet this challenge, the designer may wish to have some sense of the number of different preference orderings that are possible, given some number of alternatives for the voters to choose among. Thus, it is natural to ask: Given a set of n alternatives, how many preference orderings are there?

This problem is relatively simple if we can assume that, in every voter's preference ordering, every alternative will be ranked as either better or worse than every other alternative: that is, no two (or more) of the alternatives will be tied in any voter's preference ordering. In other words, we assume that no voter is *indifferent* between any two (or more) of the alternatives. It is well known that if we make this assumption, then a set of n alternatives gives rise to $n!$ preference orderings, since the voter can choose any of the n alternatives as his or her first choice, any of the remaining $n - 1$ alternatives as his or her second choice, and so on, down to having just 1 alternative remaining as his or her nth choice.

But a voting system must accommodate the fact that a voter may well have two (or more) alternatives tied for first place, or tied at some place further down in his or her preference ordering. In other words, a voting system must accommodate indifference. And when indifference is permitted, the number of preference orderings increases dramatically. But whereas the formula $n!$ is obvious, the analogous formula that allows for indifference is less well known. In this paper, I briefly review the prevalent approach to this problem, and then explain a less familiar approach that has certain advantages over the prevalent one.

2. *The prevalent approach: permutations of partitions*

The prevalent approach to this problem (see, e.g., [1, p. 65]) is based on the fact that the construction of a preference ordering from a set of n alternatives can be understood as a sequence of two decisions: a decision about which subsets to partition the n alternatives into, followed by a decision about the order in which to put those subsets. In effect, this approach involves grouping any alternatives that are equally good (thereby determining how many ‘levels’ the preference ordering is going to have), then placing the groups in order from best to worst. (When we do the second step, of putting the groups into some order, no group can be tied with any other. Indifference was supposed to be fully accounted for in the first step.) The alternatives are partitioned, and then the partitions are permuted.

Let us ascertain how many ways of doing this there are. As mentioned, we start with a decision about which subsets to partition the alternatives into. This perforce implies a decision about *how many* subsets to partition the alternatives into. Conveniently, this concept is captured by the Stirling numbers of the second kind. In particular, $S(n, k)$ is the number of ways of partitioning n alternatives into k subsets. Below, we will bring in the formula for $S(n, k)$, but for now let us leave $S(n, k)$ unanalysed.

Once we have selected some number of subsets k into which to partition the n alternatives, we know that we have $S(n, k)$ partitionings to choose among. We also know that regardless of which of the partitionings we choose, k subsets will result, which follows from the meaning of k, after all, and we have to put them into some order or other. Since we can choose any of the k subsets to be the first group, any of the remaining $k - 1$ subsets to be the second group, any of the remaining $k - 2$ subsets to be the third group, and so on, the subsets can be ordered in any of $k!$ ways. Thus, once we have decided to partition the n alternatives into k subsets, there are $k!\,S(n, k)$ possible preference orderings that could result.

Now, in choosing the value of k, we can choose any value from 1 to n. So, the number of possible preference orderings is $k!\,S(n, k)$ when $k = 1$, plus $k!\,S(n, k)$ when $k = 2$, plus all the values of $k!\,S(n, k)$ up to $k = n$. Thus, letting $f(n)$ be the number of preference orderings when there are n alternatives, we have the following formula for $f(n)$:

$$\sum_{k=1}^{n} k!\,S(n, k). \qquad (1)$$

Finally, let us eliminate the explicit reference to the Stirling numbers of the second kind by bringing in the formula for $S(n, k)$:

$$\frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n.$$

(See [2, p. 43]. Also see [3]. For explanation of the $\binom{k}{i}$ notation, see [2, p. 4] or [4].) Substituting this formula for $S(n, k)$ in formula (1), we have the following:

$$\sum_{k=1}^{n}k!\,\frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n.$$

This obviously simplifies to the following:

$$\sum_{k=1}^{n}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n. \tag{2}$$

This is the formula for $f(n)$ that follows from the partitioning approach to this problem.

This approach has several merits: it is logically sound; it is based on an intuitive way of addressing the problem (i.e. grouping the alternatives and then ordering the groups); and it yields a formula whose operations are agreeably elementary, involving nothing more advanced than the choose function (itself easily reduced to factorials, of course). Moreover, further work has elucidated the formula's asymptotic equivalent [5], and the error estimate involved in it [6].

This approach, however, is not the only fruitful way of addressing this problem, and I turn now to a less familiar approach. I explain this approach by examining the cases in which n equals 1, 2, and 3, then generalising to the case of an arbitrary value of n, and finally illustrating the application of the general formula to the case $n = 4$.

3. *The recursive approach suggested and derived*

Starting with the case of one alternative, we immediately find that this case is trivial. When there is just one alternative, there is just one preference ordering:

a

TABLE 1

So, we have $f(1) = 1$.

The case of two alternatives is also pretty simple. But for reasons that we will consider later, we will here break it down into two steps. First, let us look at the preference orderings that we can make if we start with just one alternative on the first level (i.e. preference orderings in which there are not any ties for first place). We can start with *a* or *b*:

a	*b*
…	…

TABLE 2

Since we have just two alternatives, each of the preference-ordering stems in this table can be completed in just one way. So we have the following preference orderings:

a	*b*
b	*a*

TABLE 3

So that's it for the preference orderings that have just one alternative on the first level. There are two of those. We also have the following preference ordering, with both alternatives on the first level:

a, *b*

TABLE 4

Obviously there is just one of those. So, when there are two alternatives, the number of preference orderings is $2 + 1$, or 3. So $f(2) = 3$.

Things get a little more complicated when there are three alternatives. First, let us look at the preference orderings that we can make if we start with just one alternative on the first level (i.e. preference orderings in which there are not any ties for first place). We can start with *a*, *b* or *c*:

a	*b*	*c*
…	…	…

TABLE 5

Let us continue each of these preference-ordering stems by putting just one alternative on the second level in each preference ordering:

a	*a*	*b*	*b*	*c*	*c*
b	*c*	*a*	*c*	*a*	*b*
…	…	…	…	…	…

TABLE 6

Since we have just three alternatives, each of the preference-ordering stems in this table can be completed in just one way. So we have the following preference orderings:

a	*a*	*b*	*b*	*c*	*c*
b	*c*	*a*	*c*	*a*	*b*
c	*b*	*c*	*a*	*b*	*a*

TABLE 7

In Table 5, we had some preference-ordering stems that we extended (in Table 6) by putting just one alternative on the second level. If we extend those stems differently, by putting two alternatives on the second level, we have the following three preference orderings:

a	*b*	*c*
b, *c*	*a*, *c*	*a*, *b*

TABLE 8

So that's it for the preference orderings that have just one alternative on the first level. There are nine of those. Now let us look at the preference orderings that we can make if we start with two alternatives on the first level:

a, *b*	*a*, *c*	*b*, *c*
…	…	…

TABLE 9

Since we have just three alternatives, each of these preference-ordering stems can be completed in just one way. So, we have the following preference orderings:

a, *b*	*a*, *c*	*b*, *c*
c	*b*	*a*

TABLE 10

And that's it for the preference orderings that have two alternatives on the first level. There are three of those. Finally, there's the preference ordering in which all three alternatives are on the first level:

a, *b*, *c*

TABLE 11

Obviously there's just one of those. So the total number of preference orderings when there are three alternatives is 9 + 3 + 1, or 13. So $f(3) = 13$.

Now, having derived this result (13 preference orderings for three alternatives) in the manner just employed, let us reconsider the case of three alternatives and observe that it can be handled in a manner that shows how we can generalise from this case to the case of an arbitrarily large number of alternatives. To begin, we found that when there are three alternatives, there are 9 preference orderings that have one alternative on the first level. We did this by exhaustively filling out and counting those preference orderings. But we could have taken a shorter (albeit less obvious) route, consisting of two steps.

The first step skips the enumeration of *a*, *b* and *c* (the three alternatives) as the ways of beginning preference orderings that have one alternative on the first level (as in Table 5) and observes that if we are going to start with one alternative on the first level, the number of ways of proceeding is the number of ways of choosing one alternative from a set of three, or $\binom{3}{1}$. (It is tempting to express this as just 3, without using combinatorial notation, but we will need the flexibility provided by combinatorial notation when we want to have more than one alternative on the first level.)

The second step skips the concrete completion of any preference-ordering stems (for there are no stems, following our new first step) and observes that if we had a preference-ordering stem with one alternative on the first level, then it could be completed by placing, below it, any of the preference orderings that consist of the remaining two alternatives. That is, the number of ways of completing this second step is $f(2)$. (Combinatorial notation, e.g.$\binom{f(2)}{1}$, is unnecessary here because we will not have occasion to imagine completing a hypothetical preference-ordering stem by choosing more than one of the ways of putting the remaining alternatives into a preference ordering.)

Since we have $\binom{3}{1}$ ways of completing the first step, and $f(2)$ ways of completing the second step, we have $\binom{3}{1}f(2)$ ways of completing the two steps together. Since $\binom{3}{1} = 3$ and we found above that $f(2) = 3$, we have 9 ways of completing this two-step process. That, of course, accords with our earlier count of 9.

Similar reasoning can account for the 3 that appears in the expression of $9 + 3 + 1$ that yields the 13 for the case of three alternatives. That 3 is the number of preference orderings that have two alternatives on the first level, and so the first step involves the number of ways of choosing two alternatives from a set of three, or $\binom{3}{2}$. The second step, in turn, involves the number of preference orderings containing the remaining alternative that can be put below the two alternatives already chosen for the first level. This number, of course, is $f(1)$. Putting these two steps together, we have $\binom{3}{2}f(1)$. Since $\binom{3}{2} = 3$ and we found above that $f(1) = 1$, we have 3 ways of completing this process. Again, this accords with our earlier count.

We have accounted for the first two summands that appear in the expression of $9 + 3 + 1$ that yields the sum of 13 for the case of three alternatives. The third summand, the 1, is fully accounted for by using just the first step of the two-step process we have been describing. For that step involves the number of ways of choosing some number of alternatives from the total set of alternatives, and the 1 corresponds to the case of putting all three alternatives on the first level. That, of course, is just $\binom{3}{3}$. Once we have chosen all three alternatives and put them on the first level, our preference ordering of three alternatives is complete. There is no remaining step corresponding to the second step of the two-step process described above.

So, the 13 we seek to explain can be written not just as $9 + 3 + 1$, but more illuminatingly as follows:

$$\binom{3}{1} f(2) + \binom{3}{2} f(1) + \binom{3}{3}.$$

Notice that each term except the final one has two factors: the first factor represents choosing some number of alternatives to put on the first level, and the second factor represents the different ways putting the remaining alternatives into a preference ordering below the first level (so that the first level and the preference ordering placed below it will together amount to a preference ordering of all of the alternatives in question). Generalising from this, it is clear that $f(n)$ can be expressed as follows:

$$\binom{n}{1} f(n-1) + \binom{n}{2} f(n-2) + \binom{n}{3} f(n-3) + \ \ldots \ + \binom{n}{n-1} f(1) + \binom{n}{n}. \quad (3)$$

This captures the idea that when there are n alternatives to be put into a preference ordering, there are n different sets of preference orderings to count up. There is the set of preference orderings that have just 1 of the n alternatives on the first level, and there are $\binom{n}{1} f(n - 1)$ of those; and there is the set of preference orderings that have 2 of the n alternatives on the first level, and there are $\binom{n}{2} f(n - 2)$ of those, and so on, through the set of preference orderings that have all n alternatives on the first level, and there are $\binom{n}{n}$ of those.

All of the terms of expression (3) could be condensed using summation notation if the final term could be made to complete the series suggested by the previous ones. Obviously what the final term needs, in order to be expressed in that way, is the additional factor of $f(0)$ after $\binom{n}{n}$. And $\binom{n}{n}$ can be rewritten as $\binom{n}{n} f(0)$ if we simply stipulate that $f(0) = 1$. So, let us rewrite expression (3) as follows:

$$\binom{n}{1}f(n-1) + \binom{n}{2}f(n-2) + \binom{n}{3}f(n-3) + \ \dots \ + \binom{n}{n-1}f(1) + \binom{n}{n}f(0), \quad (4)$$

where $f(0) = 1$. Now, condensing expression (4) using summation notation, we have the following formula for $f(n)$:

$$\sum_{i=1}^{n} \binom{n}{i} f(n - i), \text{ where } f(0) = 1. \qquad (5)$$

4. *Application to four alternatives*

To illustrate the operation of this formula, let us apply it to a slightly more complex case than the one from which we derived it. When we have four alternatives, expression (5) becomes the following:

$$\sum_{i=1}^{4} \binom{4}{i} f(4 - i), \text{ where } f(0) = 1.$$

And this, in turn, is equal to the following:

$$\binom{4}{1}f(3) + \binom{4}{2}f(2) + \binom{4}{3}f(1) + \binom{4}{4}f(0), \text{ where } f(0) = 1.$$

Implementing the stipulation that $f(0) = 1$, we have the following:

$$\binom{4}{1}f(3) + \binom{4}{2}f(2) + \binom{4}{3}f(1) + \binom{4}{4}.$$

Let us quickly evaluate and interpret each of these four terms:

- The first, $\binom{4}{1}f(3)$, is 52, and reflects the fact that if we have four alternatives and can choose one to put on the first level, then we have $\binom{4}{1}$, or 4, ways of doing that, followed by $f(3)$, or 13, ways of putting the remaining three alternatives into a preference ordering below the alternative on the first level.
- The second term, $\binom{4}{2}f(2)$, is 18, and reflects the fact that if we have four alternatives and can choose two to put on the first level, then we have $\binom{4}{2}$, or 6, ways of doing that, followed by $f(2)$, or 3, ways of putting the remaining two alternatives into a preference ordering below the two on the first level.
- The third term, $\binom{4}{3}f(1)$, is 4, and reflects the fact that if we have four alternatives and can choose three to put on the first level,

then we have $\binom{4}{3}$, or 4, ways of doing that, followed by $f(1)$, or 1, way of putting the remaining alternative into a preference ordering below the three on the first level.

- Finally, the fourth term, $\binom{4}{4}$, is 1, and reflects the fact that there is just one preference ordering in which all four alternatives are put on the first level.

So $f(4) = 52 + 13 + 4 + 1$, or 75.

5. *The recursive approach vs. the partitioning approach*

Recall the two formulas for the number of preference orderings derived in the previous sections:

$$\sum_{k=1}^{n} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k - i)^n, \tag{2}$$

$$\sum_{i=1}^{n} \binom{n}{i} f(n - i), \text{ where } f(0) = 1. \tag{5}$$

Expression (2) is the formula provided by the partitioning approach, while expression (5) is the formula provided by the recursive approach explained above.

As we noted in introducing the partitioning approach, it is the prevalent approach; it can be found in virtually every standard combinatorics textbook. Sometimes the recursive formula is listed as being equivalent to the partitioning one, but it is typically inferred from the partitioning formula rather than derived directly as in section 3 above. Thus, even when the recursive approach is presented, there is often little attention given to its underlying method (of putting some alternatives on the first level and then putting the remaining alternatives into a preference ordering below the first level).

The sequence of numbers generated by expressions (2) and (5) is the subject of [7]. There, expression (5) is preceded by ‘E.g.f.: 1/(2-exp(x))’, suggesting that it is derived from the exponential generating function, derived, in turn, from the partitioning formula, presumably. The recursive formula is also listed as being equivalent to the partitioning one (for the related problem of the number of *unordered* partitions) in some earlier works; see [8, p. xxii]; and [9, p. 210] (citing [8]). The recursive formula is also mentioned, but not explained, in [10, p. 40].

Might this relative neglect of the recursive approach be warranted? After all, as we noted after reviewing the partitioning approach, this approach has several merits, and of course in most situations, non-recursive formulas are decidedly preferable to recursive ones. However, several considerations suggest that the recursive approach may well deserve equal standing with the partitioning one.

First, expression (5) shares many of the merits mentioned in connection with expression (2) at the end of section 2. For example, in addition to being logically sound, it is based on an intuitive way of addressing the problem (i.e. building longer preference orderings from shorter ones) and it yields a formula containing no operation more advanced than the choose function.

Second, although recursive expressions often require lengthier calculations than non-recursive ones (due to the necessity of calculating the values that precede the desired one), the reverse is actually the case here. When expression (2) is used, the number of terms to be computed and summed can be counted as follows:

k	i	terms
1	0,1	2
2	0,1,2	3
3	0,1,2,3	4
…	…	…
$n-1$	$0,1,2,3, \ldots, n-1$	n
n	$0,1,2,3, \ldots, n-1, n$	$n+1$

TABLE 12

The number of terms is the sum of the numbers in the last column of this table. Because $1 + 2 + 3 + \ldots + (n-1) + n = \frac{1}{2}\left(n^2 + 1\right)$, the sum of the numbers just mentioned is $\frac{1}{2}\left(n^2 + n\right) + n$. When expression (5) is used, the number of terms can be counted as follows: $f(1)$ is just 1 (but let us regard this as a term, for counting purposes); then, with that computed, it takes 2 terms to compute $f(2)$; then, with that computed, it takes 3 terms to compute $f(3)$, and so on. So, computing $f(n)$ from scratch can be regarded as requiring $1 + 2 + 3 + \ldots + (n-1) + n$ terms. As we just saw, this series is known to sum to $\frac{1}{2}\left(n^2 + n\right)$. So, expression (5) arguably involves n fewer terms than expression (2). Admittedly, this advantage might be regarded as cancelled by the extra record-keeping demands of expression (5): rather than just maintaining a running total, as with expression (2), a user of expression (5) must keep a list of already-computed values of $f(\bullet)$. Nevertheless, at the very least, the recursive approach would appear to be on a par with the partitioning one, in terms of computational efficiency.

Third, although this record-keeping consideration complicates the recursive approach's computations in one way, it has another aspect that redounds to the distinct advantage of the recursive approach. When a task requires calculating $f(n)$ for multiple values of n, expression (5) enjoys a marked advantage over expression (2). When $f(n)$ is being calculated for some n, values of $f(\bullet)$ for numbers smaller than n have no role to play in expression (2). But in expression (5), those values of $f(\bullet)$ for numbers smaller than n can be inserted when needed in the calculation of $f(n)$. To illustrate this advantage, consider calculating $f(n)$ for $n = 1$ to $n = 10$. Using expression (2), $f(1)$ would require 2 terms, then $f(2)$ would require 5 terms, then $f(3)$ would require 9 terms, and so on, until $f(10)$ requiring 65 terms. The total number of terms is $\sum_{n=1}^{10} \frac{1}{2}\left(n^2 + n\right) + n$, or 275. In contrast, when expression (5) is used, its record-keeping component means that the total number of terms for $n = 1$ to $n = 10$ is just the number of terms for $f(10)$, which, according to the formula $\frac{1}{2}\left(n^2 + n\right)$ from the last paragraph, is just 55. In this respect, the recursive approach is clearly more computationally efficient than the partitioning approach.

Fourth, the intermediate sums generated in the course of using expression (5) are at least as meaningful as those generated in the course of using expression (2). As we saw in section 2, using expression (2) involves summing the number of ways of putting n alternatives into a preference ordering with one level ($k = 1$), the number of ways of putting n alternatives into a preference ordering with two levels ($k = 2$), and so on, up to the number of ways of putting n alternatives into a preference ordering with n levels ($k = n$). This information is meaningful, but the information involved in expression (5) is equally meaningful: every time this expression is used to calculate $f(n)$, the user recapitulates the values of $f(\bullet)$ from 1 to $n - 1$. Surely an illuminating way of arriving at the value of $f(n)$ for some n is to see how that value is based on the values of $f(n)$ for all the smaller values of n.

Overall, the recursive approach has much to recommend it. The partitioning approach has value as well, but the recursive approach offers a different perspective that has distinct strengths in terms of both the efficiency and the meaningfulness of the calculations it requires. It is an important approach to solving the problem of the number of preference orderings.

Acknowledgements

I would like to thank my father, Joe Eggleston, for helping me with my initial explorations of the problem this article discusses. I would also like to thank Jeff VanderKam, Rodolfo Torres, and especially Jeremy Martin for sharing their expertise and directing me to several pertinent combinatorics resources. Finally, I would like to thank the anonymous referee for this journal whose suggestions led to several stylistic improvements.

References

1. J. L. Chandon, J. Lemaire, and J. Pouget, Dénombrement des quasi-ordres sur un ensemble fini, *Mathématiques et Sciences Humaines* **62** (1978) pp. 61-80.
2. J. Riordan, *An introduction to combinatorial analysis*, John Wiley & Sons, Inc., New York (1958).

3. E. W. Weisstein, Stirling number of the second kind, *MathWorld*, accessed December 2014 at http://mathworld.wolfram.com/StirlingNumberoftheSecondKind.html

4. E. W. Weisstein, Binomial coefficient, *MathWorld*, accessed December 2014 at http://mathworld.wolfram.com/BinomialCoefficient.html

5. J. P. Barthelemy, An asymptotic equivalent for the number of total preorders on a finite set, *Discrete Mathematics* **29** (1980) pp. 311-313.

6. R. W. Bailey, The number of weak orderings of a finite set, *Social Choice and Welfare* **15** (1998) pp. 559-562.

7. N. J. A. Sloane, *Online encyclopedia of integer sequences*, accessed December 2014 at http://weis.org/A000670

8. A. C. Aitken, A problem in combinations, *Edinburgh Mathematical Notes* **28** (1933) pp. xviii-xxiii.

9. L. Comtet, *Advanced combinatorics: the art of finite and infinite expansions* (revised edition), D. Reidel Publishing Company, Dordrecht, Holland (1974).

10. G. Kreweras, Une dualité élémentaire souvent utile dans les problèmes combinatoires, *Mathématiques et Sciences Humaines* **3** (1963) pp. 31-41.

doi: 10.1017/mag.2014.3

BEN EGGLESTON

Department of Philosophy, University of Kansas, Lawrence, KS 66045, USA

e-mail: *eggleston@ku.edu*