----
4. EXPOSITORY ARTICLE

DISCRETE MATHEMATICS AND
COUNTING DERANGEMENTS IN BLIND WINE TASTINGS

JOHN D. NYSTUEN
College of Architecture and Urban Planning
The University of Michigan

SANDRA L. ARLINGHAUS
School of Natural Resources and Environment
The University of Michigan

WILLIAM C. ARLINGHAUS
Department of Mathematics and Computer Science
Lawrence Technological University

The statistician Fisher explained the mathematical basis for the field of "Design of Experiments" in an elegant essay couched in the context of the mathematics of a Lady tasting tea (Fisher, in Newman 1956; Fisher 1971). In Fisher's text, the problem is to analyze completely the likelihood that the Lady can determine whether milk was added to the tea or tea added to milk. Problems associated with the tasting of wines have a number of obvious similarities to Fisher's tea-tasting scenario. We offer an analysis of this related problem, set in the context of Nystuen's wine tasting club. To begin, a brief background of the rules of that club seems in order; indeed, it is often the case that the application is forced to fit the mathematics in order to illustrate the abstract. Here, it is the real-world context that guides the mathematics selected.

Wine Tasting Strategy

The Grand Crew wine club of Ann Arbor has been blind-tasting wines monthly for years. In a blind tasting, several wines are offered with their identity hidden. Not only are labels covered, but the entire bottle is covered as well because the shape and color of the bottle provides some clues as to the identity of the wine. The wines are labeled 1 through n in the order presented. Six to eight wines are tasted at a sitting. Members sip the wines and score each on a scale from 1 to 20, using a scoring method suggested by the American

Wine Association. The wines are judged on the basis of quality and individual taster preference. The evening's host is in charge of choosing and presenting the wines. Usually wines of a single variety but from different vineyards, wineries, prices, or distributors are tasted. Two sheets of paper are provided to each taster. One is a blank table with a row for each wine numbered 1 through n in the order presented. The columns on this sheet provide space for comments, the individual's numerical ratings of the wines, average ratings of the group, and the range in scores for each wine. One column is reserved for the member's guess as to the identity of the wine. The second sheet contains information about each wine to be used to match the wines tasted. On this sheet the wines are labeled a, b, c, and so forth, along with information on age, winery, negotiant, and price. The tasters try to match the identity of the wine with their individual rating on sheet 1.

The wines are listed in unknown order on the second sheet. The tasters make their decisions by matching the letter identification with the numerical order of presentation. On rare occasions one or more members correctly identifies every wine. More often two or more wines are mislabeled, and quite often the identities seem hopelessly scrambled. Guessing at random would seem just as effective. The question then arises, "what are the chances of getting one, two, more, or all correct by chance alone?" Discrete mathematics and the algebra of derangements provides the answer to this question.

Probabilities are a matter of counting. In what proportion does a particular combination of correct and incorrect identifications occur purely at random out of all possible combinations? The denominator in this proportion is a count of all possible arrangements and the numerator is a count of all possible ways a particular event occurs, such as one right, all the rest wrong. The denominator is easily determined. If one has five things any of the five might be chosen first; there remain four things any of which might be chosen next. The process continues until the last stage in which only one can be chosen. Thus, there are 5*4*3*2*1=120 ways to arrange five bottles of wine--the customary notation for this product is 5! (read five factorial). This notation extends to arbitrarily large positive integers in the obvious way; 0! is defined to be 1. The factorial of a number grows rapidly with an increase in the size of the number; thus, 7!=5040 while 8!=40,320.

The numerator of the proportion sought is not found as easily. Consider the case of a blind tasting of three bottles of wine. Suppose the first one is correctly identified; the remaining two outcomes must be both right or both wrong. It is not possible to identify two wines correctly and the third one incorrectly. Table 1 illustrates all possible patterns of identification for three bottles of wine, a, b, and c, with bottle "a" presented first, bottle "b" presented second, and bottle "c" presented third. As this table indicates, there is only one arrangement in which all are correct, two arrangements with none correct and three arrangements with two correct. There are, of course, no arrangements with exactly one correct.

Table 1. All possible arrangements of three items, a, b, and c.
            Number of matches and non-matches to the arrangement abc.

Matches		Non-matches
abc	3	0
acb	1	2
bac	1	2
bca	0	3
cab	0	3
cba	1	2

When all possible outcomes, shown in Table 1, are enumerated, it is an easy matter to calculate the probability of each type of event-- to obtain the probability, divide each outcome from Table 1 by 3!, the number of total possible arrangements. Table 2 shows the probability of each outcome: P(0) denotes none right, P(1) denotes exactly one right, and so forth. The sum of all probabilities adds to 1.00, as it should.

Table 2. Probability of a correct labeling.

$P(0) = 2/6 = 0.33$
$P(1) = 3/6 = 0.50$
$P(2) = 0/6 = 0.00$
$P(3) = 1/6 = 0.17$

A total enumeration approach to finding the probabilities is satisfactory for introductory purposes and for very small samples. Even for six, seven, or eight wines at a single tasting it is, however, not satisfactory; Table 1 would expand to 720, 5040, or 40,320 columns for each of those cases. Clearly more clever and mathematically elegant ways of counting, rather than brute force listings, are required. In this latter regard, one is reminded of the story of Gauss who, as a young child, astounded his German schoolteacher with an instant result for what the teacher had planned as a tedious exercise. The teacher, in order to keep his students busy, told them to add all the numbers from 1 to 100. Gauss immediately wrote the answer on his slate. He had apparently discovered for himself that the sum, S, of the first n positive integers is given by the recursive relationship $S=(n(n+1)/2)$. Thus, all he had to do was multiply 50 by 101 to obtain the answer: an elegant solution to an otherwise tedious problem. It was the more mature Gauss and later Laplace that would do pioneering work in the Theory of Errors of Observation which in turn would serve as a significant part of the base for applications of mathematics and statistics (in Design of Experiments) in the Scientific Method.

Derangements

For our problem, we need a way to count the number of times a taster can get all the wines right, one wine right and all the others wrong, two wines right and all the others wrong, and so forth. To convert the tedious, brute force task of listing permutations and combinations for this problem, to a more tractable situation, we employ the concept of "derangement," that will eliminate, notationally, combinations that we do not wish to consider.
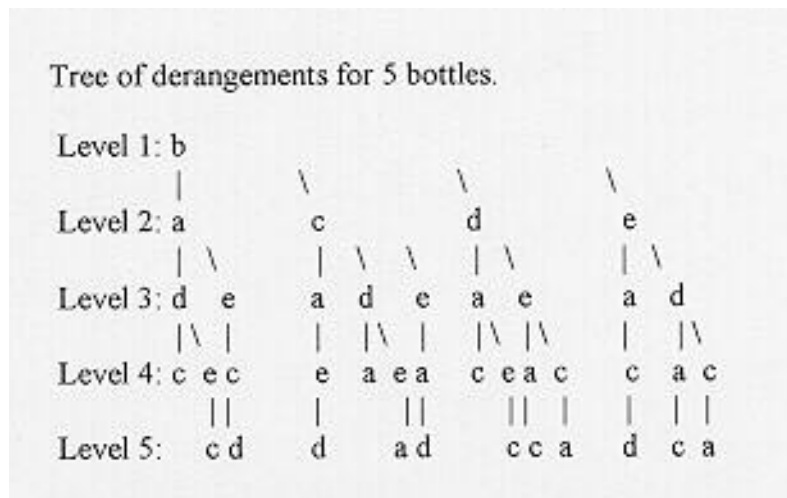
A "derangement" is a permutation of objects that leaves no object in its original position (Rosen 1986; Michaels and Rosen 1991). The permutation badec is a derangement of abcde because no letter is left in its original position. However, baedc is not a derangement of abcde because this permutation leaves d fixed. Thus, the number of times a wine taster gets all the wrong answers in tasting n bottles is the number of derangements of n numbers, D(n), divided by n!: D(n)/n!. The value of D(n) is calculated as a product of n! and a series of terms of alternating plus and minus signs:

$D(n)=n!(1-1/1!+1/2!-1/3!+1/4!-1/5!+...+((-1)^n)/n!).$

Readers wishing more detail concerning this formula might refer to Rosen (1988); for the present, we continue to consider the use of derangements.

In order to see how derangements can be enumerated visually, we construct the following tree of possibilities for arrangements of 5 letters which do not match the natural order of abcde. On the first level, the natural choice is a--so choose some other letter instead. The second level would be b in the natural order so choose all others, instead, and continue the process until all possibilities have been exhausted. Following each path through the tree will give all possible derangements beginning with the letter b--there are 11 such routes. Thus, there are 11*4 derangements.

Tree of derangements for 5 bottles.

Tree of derangements for 5 bottles.

```
Level 1: b
         |      \            \           \
Level 2: a       c            d           e
         | \     | \  \       | \         | \
Level 3: d  e    a  d  e      a  e        a  d
         |\ |    |  |\ |      |\ |\       |  |\
Level 4: c e c   e  a e a     c ea c      c  a c
         ||      |    ||      || |        |  | |
Level 5: c d     d    a d      c c a      d  c a
```

Indeed, when there are five wines, $D(n)=5!(1/2!-1/3!+1/4!-1/5!)=5!/2!-5!/3!+5!/4!-5!/5!=60-20+5-1=44$.

What is of particular significance is that derangements focus only on wrong guesses: because a non-wrong guess is a correct guess, it is possible to focus only on one world. The Law of the Excluded Middle, in which any statement is "true" or "false"--with no middle partial truth admitted, is the basis for this and for most mathematical assessments

of real-world situations. It is therefore important to use the tools appropriately, on segments of the real-world situation in which one can discern "black" from "white."

Derangements and Probability in Random Guesses

In the case of the five wine example, the number of ways of choosing (for example) three correctly out of five is the combination of five things taken three at a time: C(5,3)=5!/2!3!=10. Exhausting all possible combinations reflects an expected connection with the binomial theorem--these values are the coefficients of (x+y)^5.

C(5,0)=1
C(5,1)=5
C(5,2)=10
C(5,3)=10
C(5,4)=5
C(5,5)=1.

The total number of right/wrong combinations is therefore 2^5 or 32. Notice, though, that the pattern within each grouping is disregarded; to discover the finer pattern, of how right/wrong guesses are arranged we need permutations. To limit the number of permutations necessary to consider, we investigate the derangements.

If we can count derangements, we can now address the question of how many times a taster, guessing randomly, gets exactly one wine correct. The answer is simply the number of ways one bottle can be chosen from n bottles times the number of derangements of the other (n-1) bottles of wine. When this value is divided by n!, the probability P(1) of guessing exactly one wine correctly is the result. That probability is:

P(1)=(n!/1!(n-1)!)*D(n-1)/n!

This idea generalizes in a natural manner so that the probability of choosing exactly k wines correctly is given as:

P(k)=(n!/k!(n-k)!)*D(n-k)/n!

Table 3 displays all the probabilities for outcomes in blind tastings in which random choices are made in situations for which from 2 to 8 wines are offered by the evening's host. Notice that there is less than a one percent chance of guessing all wines correctly by chance alone whenever the host offers five or more wines in the evening's selection. Evidently, some knowledge of wines is displayed by a taster who accomplishes this feat

with any regularity. On the other hand, one could expect, by chance alone, to guess none of the wines correctly about 37 percent of the time, independent of the number of wines offered for tasting. The same situation holds for guessing exactly one wine correctly.

The reason that this is so, as readers familiar with infinite series will note, is that the alternating series contained in the parenthetical expression in the formula for counting derangements is precisely 1/e, where e is the base of natural logarithms (a transcendental number of value approximately 2.71828). That is, $e^x = 1+x/1!+(x^2)/2!+(x^3)/3!+...$ so that when x=-1, then $e^{(-1)}$, or 1/e, is precisely the parenthetical expression in the formula for D(n). The larger the value of n, the closer the approximation to 1/e=0.3678797. In a blind tasting with an infinite number of bottles of wine, random choices will result in approximately a 0.368 probability that all will be in error!

Table 3. Probability of correctly matching K wines from tasting a total of n wines

n

K	2	3	4	5	6	7	8
0	.5000	.3333	.3750	.3667	.3681	.3679	.3679
1	.0000	.5000	.3333	.3750	.3667	.3681	.3679
2		.0000	.2500	.1667	.1875	.1833	.1840
3		.1667	.0000	.0833	.0555	.0625	.0611
4			.0417	.0000	.0208	.0139	.0156
5				.0083	.0000	.0042	.0028
6					.0014	.0000	.0007
7						.0002	.0000
8							.0000

Table 3 suggests some rules of thumb about how well a taster has done. In a normal-sized tasting of six, seven, or eight wines, identifying at least five of them correctly occurs less than 1% by chance alone. Identifying four correctly happens by chance about 2 percent (or less) of the time. However, identifying three correctly occurs by chance from near five to six percent of the time: in every 16 to 18 tastings. Usually there are ten to twelve tasters at a sitting in this one club. None to one member at a sitting rates to guess three

wines correctly by chance alone; the group usually does substantially better than this, suggesting some expertise in identifying the wines.

The Principle of Inclusion and Exclusion: The Basis for Counting.

The expression for counting derangements, as a product of n! and and a truncated series for 1/e, has some interesting properties, most notably perhaps the alternating plus and minus signs preceding terms of the series. This alternation occurs because the principle of inclusion and exclusion has been used as the basis for the counting.

Readers versed in elementary set theory, Boolean algebra, or symbolic logic, are familiar with the idea of including the intersection, and then subtracting it out, in order to count the number of elements in intersecting sets. This idea, in this context, was clearly familiar to Augustus DeMorgan in the late nineteenth century. Indeed, in a wider context, it dates back to the time of Eratosthenes of Alexandria and his sieve for determining which numbers are prime: those that are multiples of numbers early in the ordering of positive integers are excluded. Only those numbers not excluded have divisors of only themselves and 1, and so are exactly the set included as prime numbers.

The following example illustrates how inclusion and exclusion is used in counting derangements; the reader interested in the general proof is referred to Rosen (1988). It is easy to visualize cases when n is small using Venn diagrams--thus, the linkage between inclusion/exclusion, set theory, and derangements becomes clear.

Consider for example, a tasting of two wines. Let a be the event that the first wine is correctly identified; let b be the event that the second wine is correctly identified. Draw a rectangle on a sheet of paper and within the rectangle draw two intersecting circles, a and b--a familiar Venn diagram. The content of the rectangle is the universe of discourse. The content of circle a is the set of all events that the first wine is correctly identified (either alone or with another), denoted N(a). The content of circle b is the set of all events that the second wine is correctly identified, denoted N(b). The intersection of the two circles has content ab, the set of all events in which both the first wine and the second wine are correctly identified, denoted N(ab). The set of all derangements is the content of that area of the rectangle outside the two circles. The content of the two circles is the sum of the

content of the first circle plus the sum of the content of the second circle: N(a)+N(b). This sum however includes N(ab) in the first term and also N(ab) in the second term; thus, N(ab) must be excluded from the sum to get an accurate count of the content of the union of the two circles--hence inclusion and exclusion. The accurate count of one or more wines correct is thus given as N(a)+N(b)-N(ab). The case for three circles is more complicated to visualize but can be enumerated carefully as a set of three two-circle problems. With values greater than 3, visualization in this manner becomes impossible and one must rely on extension of the notation and visualization in the world of language rather than in the world of pictures--both subsets of "the world of mathematics." Indeed, geographers interested in spatial statistics should be familiar with this issue in using the statistical forms to capture what becomes increasingly too complex to map.

Retrospect

   These classical ideas, whether cast in the number theoretic context of prime numbers, in the discrete mathematics context of inclusion and exclusion, or in the set theoretic context of intersections, served once again, when cast in the context of derangements and the counting of incorrectness, to permit a clever solution to a complicated, uncontrived, real-world problem. What this sort of analysis offers is a challenge to look at the world in different ways: from the use of classical theoretical material in new real world situations, to the development of new theoretical material which can foster further theoretical exploration and application.

References

Fisher, R. A. The Design of Experiments. Eighth edition, reprinted,   New York, Hafner and Co., 1971. First edition, Edinburgh, London, Oliver and Boyd, 1935.

Fisher, R. A. "The Mathematics of a Lady Tasting Tea" in Newman, J.R., ed. The World of Mathematics, Simon and Shuster, New York, 1956 (pp. 1512-1521).

Michaels, John G. and Rosen, Kenneth H. (eds.) Applications of Discrete Mathematics, New York, McGraw-Hill, 1991.

Polya, G.; Tarjan, R. E.; and, Woods, D.R.   Notes on Introductory Combinatorics, Boston, Birkhauser, 1983.

Rosen, Kenneth R. Discrete Mathematics and Its Applications. First edition, New York, Random House, 1988.