The Fourth Way to Sample *k* Obejcts from a Collection of *n*

By Kurt Kreith and Judith Kysh, University of California, Davis, CA 95616

As probability and finite mathematics find their way into growing numbers of mathematics classrooms, the ideas underlying combinatorial analysis will become increasingly important in the secondary school curriculum. The challenge here is not just to teach the various counting rules and formulas for permutations and combinations but also to supply the student with a means for deciding when and where these concepts are appropriate to apply.

To help with this decision-making dilemma, it is instructive to consider all possibilities for "sampling a collection of *n* objects *k* times." Before one can determine whether such a sampling process leads to permutations, combinations, or some other combinatorial rule, it is necessary to specify two aspects of the sampling procedure. In particular,

1. are we allowed to sample the same objects more than once? And
2. is the order in which our sample is drawn significant, or just of the sample itself?

Since each of these questions admits two answers, common sense (or elementary combinatorial analysis) makes clear that sampling *k* objects from a collection of *n* can be done in *four* different ways. Three of these ways lead to familiar combinatorial concepts. However, the fourth is often overlooked, even though it is dealt with in some standard treatises on probability theory (Feller 1957, 67-70; Parzen 1960, 38-41.) It has an interesting history, and its inclusion in the introductory combinatorics curriculum could help students to organize their decision making when categorizing counting problems.

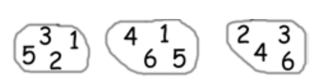
Some Familiar Concepts

To be Specific, let's assume that $n = 6$ and $k = 4$. If we choose to sample with repetition (think of replacing the sampled object after it has be chosen) and we distinguish samples by order, then this sampling process can be simulated by rolling a six-sided die four times and recording the sequences of outcomes. We could enumerate all possible outcomes in an ordered listing such as (1,1,1,1), (1,1,1,2), (1,1,2,1),…,(5,3,3,5),…,(6,6,6,6). The "fundamental principle of counting" or "multiplication rule" applied to this situation leads to $6 \times 6 \times 6 \times 6 = 6^4$ possible outcomes.

Suppose that we now specify that the sampling be done without repetition (replacement.) In this instance we can think of a bag containing six tiles labeled 1, 2, 3, 4, 5, and 6. We pull out 4 of the 6 tiles from the bag and place them in order. We could enumerate all possible outcomes in an ordered listing such as (1,2,3,4), (1,2,3,5), (1,2,3,6), (1,2,4,3), …, (4,3,2,1), …, (6,5,4,3). Again, we use the counting principle to conclude there are $6 \times 5 \times 4 \times 3 = 360$ different outcomes. As such, the process of sampling without repetition (replacement), but with respect to order, leads to the familiar concept of permutations and the symbol

$${}_6P_4 = 6 \times 5 \times 4 \times 3 = \frac{6!}{2!}.$$

Thirdly, let us suppose that we are sampling without replacement and that order is not significant. Here we can think of the same six tiles in a bag and choosing a sample of four tiles all at once in a jumble. Such a process will lead to outcomes that can be illustrated by sets like so:



where the irregular shapes are meant to stress that order is irrelevant. Since each of these unordered sets can be ordered in 4! different ways, we can compute the number of different unordered combinations by dividing the number of permutations, ${}_6P_4$ by 4!: ${}_6C_4 = \frac{{}_6P_4}{4!} = \frac{6!}{2!4!}$.

The Fourth Way

We can summarize our results so far in a table in which we generalize the specific values 6 and 4 in the previous examples by *n*: the number of objects to sample from and *k*: the number of samples.

TABLE 1			
Replacement allowed?	Order matters?	Number of outcomes	Symbol
yes	yes	n^k	None
no	yes	$\frac{n!}{(n-k)!}$	${}_nP_k$
no	no	$\frac{n!}{k!(n-k)!}$	${}_nC_k$

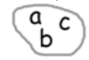
The sampling procedure that has been missed is clearly one that allows replacement but is regardless of order. This sampling procedure can be simulated by simultaneously rolling $k = 4$ indistinguishable dice with $n = 6$ numbered sides. In this situation, the outcomes (2,2,4,5) and (5,2,4,2) are counted as the same.

Before continuing with the computation of the number of outcomes when *k* *n*-sided dice are rolled, let's digress with a discussion of the history of such counting problems.

Historical Observations

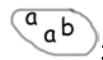
In her book *Games, Gods and Gambling*, David (1962) calls attention to a moral dice game invented by Bishop Wibold of Cambrai during the tenth century A.D. This game called for the player to roll three six-sided dice of the same color (indistinguishable) and associated a virtue with each outcome.
The problem of enumerating the set of outcomes was sufficiently challenging as to be the subject of a poem entitled "De Vetula," which was published in Latin during the thirteenth century. Whereas in the case of $n = 6$ and $k = 3$ one can readily enumerate the 56 different outcomes (virtues?), the case $n = 6$ and $k = 4$ already begins to suggest the need for a general formula.

In the context of probabilities, another important observation arises regarding the roll of three dice. Assuming the dice are fair, we must keep track of order if we are to generate a sample space of equally likely outcomes. In the example at hand we are, by ignoring order, collapsing a sample space of $6 \times 6 \times 6 = 216$ equally likely outcomes into one with 56. That is, for example, the six equally likely outcomes, *abc*, *acb*, *bac*, *bca*, *cab*, *cba* collapse into a set:



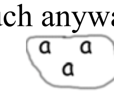
;

The three equally likely outcomes *aab*, *aba*, *baa* collapse into a set:



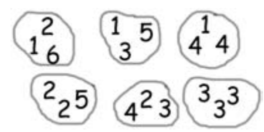
;

and, of course, in the single outcome, aaa, order doesn't matter much anyway:

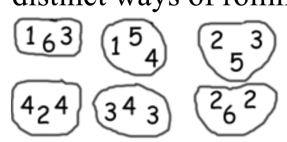


.

These considerations arose during the early seventeenth century when the Grand Duke of Tuscany posed to Galileo a famous problem regarding the likelihood of rolling a 9 versus a 10 with three dice. The duke had observed that although six visually distinct ways exist of rolling a 9,



and six visually distinct ways of rolling a 10,



his experiences suggested that a 10 is more likely than a 9. Galileo's response was, in essence, to point out that the duke's sample space of 56 outcomes is not simple (i.e., not composed of equally likely outcomes.) In this way Galileo was able to show that

$$P(9) = \frac{25}{216}$$

whereas

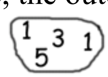
$$P(10) = \frac{27}{216}$$

corroborating the duke's impressions based on extensive dicing.

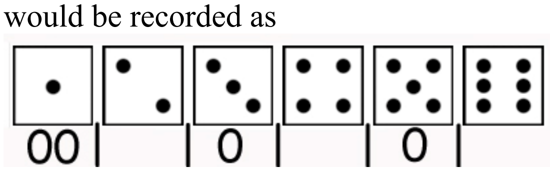
The General Case

In honor of Bishop Wibold, we denote by ${}_nW_k$ the number of visually distinct outcomes that can occur when a set of *n* objects is sampled *k* times with replacement (repetition) but without regard to order—in other words, the number of visually distinct outcomes that can occur in the roll of *k* dice, each of which has *n* faces. It has already been noted that ${}_6W_3 = 56$ was established over 700 years ago, but a corresponding attempt to find ${}_6W_4$ by direct enumeration would be considerably more tedious. However, a simple way of calculating ${}_6W_4$, and more generally ${}_nW_k$, exists for arbitrary values of *n* and *k*.

Motivated by Feller (1957), we begin with a graphical technique for recording the visually distinct outcomes corresponding to the roll of *k* identical dice. Given a horizontal arrangement of pictures for the six faces of a standard die, one can record the number of times a particular face comes up by place that number of zeros underneath the corresponding picture. Thus in a roll of $k = 4$ dice, the outcome



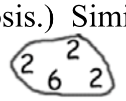
would be recorded as



Notice that we have separated the zeros corresponding to each face of the die by a vertical bar. The bar helps to ekep the notation precise, and it also serves another important function. Specifically, inserting the bars allows us to erase the pictures of the faces of the die and to record the outcomes in the form

0 0 1 1 0 1 1 0 1

(in which the vertical bars have undergone a slight metamorphosis.) Similarly the roll



would now be recorded as

1 0 0 0 1 1 1 1 0

Recall now that each linear ordering of *k* zeros and *n* – *k* ones corresponds to choosing the *k* spaces to be occupied by zeros (or the *n* – *k* spaces to be occupied by ones.) Since the notation establishes a 1-1 correspondence between the visually distinct outcomes possible in the roll of four identical dice and all possible sequences of four zeros and five ones, it follows that ${}_6W_4 = {}_{5+4}C_4 = 126$, while more generally,

$${}_6W_k = {}_{5+k}C_k = \frac{(5+k)!}{5!k!}.$$

Repeating the same reasoning for *k* identical dice with *n* faces, we obtain the general formula

$${}_nW_k = {}_{n-1+k}C_k = \frac{(n-1+k)!}{(n-1)!k!}$$

By way of a check, it is reassuring to note that this yields

$${}_6W_3 = {}_8C_3 = \frac{8!}{5!3!} = 56$$

As enumerated in “De Vetula.”
Now we can complete the table previously considered:

TABLE 2			
Replacement allowed?	Order matters?	Number of outcomes	Symbol
yes	yes	n^k	None
no	yes	$\frac{n!}{(n-k)!}$	${}_nP_k$
yes	no	$\frac{(n-1+k)!}{(n-1)!k!}$	${}_nW_k$
no	no	$\frac{n!}{k!(n-k)!}$	${}_nC_k$

Although completing the set of alternative types of counting problems complicates matters with the development of another formula, the student gains the advantage of being able to classify all sampling problems as determined by order and repetition. It is hoped that this process of classification will also furnish a basis for approaching such problems with increased confidence and success.

REFERENCES

David, Florence N. Games, Gods and Gambling. New York: Hafner, 1962

Feller, William. An Introduction to Probability theory and Its Applications. Vol. 1. New York: John Wiley and Sons, 1957

Parzen, Emmanuel. Modern Probability Theory and Its Applications. New YorkL John Wiley and Sons. 1960.