



2.1.4 Unordered Sampling with Replacement

Among the four possibilities we listed for ordered/unordered sampling with/without replacement, unordered sampling with replacement is the most challenging one.

Suppose that we want to sample from the set $A = \{a_1, a_2, \dots, a_n\}$ k times such that repetition is allowed and ordering does not matter. For example, if $A = \{1, 2, 3\}$ and $k = 2$, then there are 6 different ways of doing this

- 1,1;
- 1,2;
- 1,3;
- 2,2;
- 2,3;
- 3,3;

How can we get the number 6 without actually listing all the possibilities? One way to think about this is to note that any of the pairs in the above list can be represented by the number of 1's, 2's and 3's it contains. That is, if x_1 is the number of ones, x_2 is the number of twos, and x_3 is the number of threes, we can equivalently represent each pair by a vector (x_1, x_2, x_3) , i.e.,

- $1,1 \rightarrow (x_1, x_2, x_3) = (2, 0, 0)$;
- $1,2 \rightarrow (x_1, x_2, x_3) = (1, 1, 0)$;
- $1,3 \rightarrow (x_1, x_2, x_3) = (1, 0, 1)$;
- $2,2 \rightarrow (x_1, x_2, x_3) = (0, 2, 0)$;
- $2,3 \rightarrow (x_1, x_2, x_3) = (0, 1, 1)$;
- $3,3 \rightarrow (x_1, x_2, x_3) = (0, 0, 2)$.

Note that here $x_i \geq 0$ are integers and $x_1 + x_2 + x_3 = 2$. Thus, we can claim that the number of ways we can sample two elements from the set $A = \{1, 2, 3\}$ such that ordering does not matter and repetition is allowed is the same as solutions to the following equation

$$x_1 + x_2 + x_3 = 2, \text{ where } x_i \in \{0, 1, 2\}.$$

This is an interesting observation and in fact using the same argument we can make the following statement for general k and n .

Lemma 2.1

The total number of distinct k samples from an n -element set such that repetition is allowed and ordering does not matter is the same as the number of distinct solutions to the equation

$$x_1 + x_2 + \dots + x_n = k, \text{ where } x_i \in \{0, 1, 2, 3, \dots\}.$$

So far we have seen the number of unordered k -samples from an n element set is the same as the number of solutions to the above equation. But how do we find the number of solutions to that equation?

Theorem 2.1

The number of distinct solutions to the equation

$$x_1 + x_2 + \dots + x_n = k, \text{ where } x_i \in \{0, 1, 2, 3, \dots\} \quad (2.3)$$

is equal to

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}.$$

Proof

Let us first define the following simple mapping in which we replace an integer $x_i \geq 0$ with x_i vertical lines, i.e.,

$$\begin{aligned} 1 &\rightarrow | \\ 2 &\rightarrow || \\ 3 &\rightarrow ||| \\ &\dots \end{aligned}$$

Now suppose we have a solution to the Equation 2.3. We can replace the x_i 's by their equivalent vertical lines. Thus, for example if we have

$x_1 + x_2 + x_3 + x_4 = 3 + 0 + 2 + 1$, we can equivalently write $||| + +|| + |$. Thus, we claim that for each solution to the Equation 2.3, we have unique representation using vertical lines (|) and plus signs (+). Indeed, each solution can be represented by k vertical lines (since the x_i sum to k) and $n - 1$ plus signs. Now, this is exactly the same as [Example 2.7](#): how many distinct sequences you can make using k vertical lines (|) and $n - 1$ plus signs (+)? The answer as we have seen is

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}.$$

Example 2.12

Ten passengers get on an airport shuttle at the airport. The shuttle has a route that includes 5 hotels, and each passenger gets off the shuttle at his/her hotel. The driver records how many passengers leave the shuttle at each hotel. How many different possibilities exist?

Solution

Let x_i be the number of passengers that get off the shuttle at Hotel i . Then we have

$$x_1 + x_2 + x_3 + x_4 + x_5 = 10, \text{ where } x_i \in \{0, 1, 2, 3, \dots\}.$$

Thus, the number of solutions is

$$\binom{5 + 10 - 1}{10} = \binom{5 + 10 - 1}{5 - 1} = \binom{14}{4}.$$

Let's summarize the formulas for the four categories of sampling. Assuming that we have a set with n elements, and we want to draw k samples from the set, then the total number of ways we can do this is given by the following table.

ordered sampling with replacement	n^k
ordered sampling without replacement	$P_k^n = \frac{n!}{(n-k)!}$
unordered sampling without replacement	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$
unordered sampling with replacement	$\binom{n+k-1}{k}$

Table 2.1: Counting results for different sampling methods.
