Combinatorial Analysis

Quasar Chunawala

Abstract

A short note on combinatorial analysis

---

1 Preliminaries.

Pairs.

With m elements $\{a_1, a_2, \ldots, a_m\}$ and n elements $\{b_1, b_2, \ldots, b_n\}$, it is possible to form mn pairs (a_i, b_j) containing one element from each group.

Proof. Arrange the pairs in a rectangular array in the form of multiplication table with m rows and n columns, so that (a_i, b_j) stands at the intersection of the ith row and the jth column. Then, each pair appears once and only once and the assertion becomes obvious.

Example 1.1. **Bridge Cards.** As sets of elements we take the four suits $\{\text{Heart}, \text{Diamond}, \text{Club}, \text{Spade}\}$ and thirteen face values $\{1, 2, 3, \ldots, 13\}$ respectively. Each card is defined by its suit and face value, and so there exist $4 \cdot 13 = 52$ such combinations or cards.

Multiplets. Given n_1 elements of $\{a_1, a_2, \ldots, a_{n_1}\}$, n_2 elements of $\{b_1, \ldots, b_{n_2}\}$ upto n_r elements of $\{x_1, x_2, \ldots, x_{n_r}\}$ it is possible to form $n_1 \cdot n_2 \cdot n_3 \cdots n_r$ ordered r-tuplets $(a_{j_1}, b_{j_2}, \ldots, x_{j_r})$ containing one element of each kind.

Perhaps the smallest and the most useful way of describing the last theorem is as follows. To form an r-tuplet, we have to choose one a, one b, etc. We have to perform r selections in all and have in succession $n_1, n_2, \ldots, n_r$ possibilities to choose from. Similar to forming pairs, it is intuitive that this procedure leads to $n_1 \cdot n_2 \cdots n_r$ different results.

Example 1.2. Multiple Classifications. Suppose that people are classified according to their marital status, sex and profession. The various categories play the role of elements. If there are 17 professions, we have $2 \cdot 2 \cdot 17 = 68$ classes in all.

Example 1.3. In an agricultural experiment three different treatments are to be tested. If these treatments can be applied on r_1, r_2 and r_3 levels of concentration, then there exist a total of r_1, r_2, r_3 combinations or way of treatment.

Example 1.4. Placing balls into cells amounts to choosing one cell for each ball. With r balls, we have r independent choices, and therefore r balls can be place into n cells in n^r different ways. It will be recalled that a great variety of conceptual experiments are abstracting equivalent to placing balls into cells.

For example, considering the faces of a die as cells, the last proposition implies that the experiment of throwing a die r times has 6^r possible outcomes of which 5^r satisfy the condition that no six turns up.

Assuming that all outcomes are equally probable, the event that no six in r throws has therefore the probability $(5/6)^r$.

2 Ordered Samples.

Consider the set or population of n elements $a_1, a_2, \ldots, a_n$. Any ordered arrangement $a_{j_1}, a_{j_2}, \ldots, a_{j_r}$ of r symbols is called an ordered sample of size r drawn from our population. For an intuitive picture, we can imagine that the elements are selected one by one. Two procedure are then possible. First sampling with replacement; here each selection is made from the entire population one by one, so that the same element can be drawn more than once. The samples are then arrangements in which repetitions are permitted.

Second, sampling without replacement; here an element once chosen is removed from the population, so that the sample becomes an arrangement without repetition. Obviously, in this case the sample size r cannot exceed the population size n.

In sampling with replacement, each of the r elements can be chosen in n ways: the number of possible samples is therefore n^r, as can be seen from the last theorem with $n_1 = n_2 = \ldots = n$. In sampling without replacement, we have n possible choices for the first element, but only $(n-1)$ for the second, $(n-2)$ for the third etc. Using the same rule, we see that in this case we have $n(n-1)\cdots(n-r+1)$ choices in all. Products of this type appear so often that it is convenient to introduce the notation

$$(n)_r = n(n-1)\cdots(n-r+1)$$

Clearly, $(n)_r = 0$ for integers $r > n$. We have thus, the following.

Theorem 2.1. For a population of n elements and a prescribed sample size r, there exist n^r different samples with replacement and $(n)_r$ samples without replacement.

We note the special case, where $r = n$. In sampling without replacement, a sample size n includes the whole population and represents a reordering (or permutation) of its elements. According n elements can be ordered in $(n)_n = n \cdot (n-1)\cdots 2 \cdot 1$ different ways. Instead of $(n)_n$, we write $n!$ which is the usual notation. We see that our theorem has the following corollary.

Corollary 2.1.1. The number of different orderings of n elements is $n! = n \cdot (n-1)\cdots 2 \cdot 1$.

Drawing r elements from a population of size n is an experiment whose possible outcomes are samples of size r. Their number is n^r or $(n)_r$ depending on whether or not replacement is used.In either case, our conceptual experiment is described by a sample space in which each individual point represents a sample of size r.

So far, we have not spoken of probabilities associated with our samples. Usually, we shall assign equal probabilities to all of them and then speak of **random samples**. The word random is not well-defined but when applied to samples or selections it has a unique meaning. Whenever we speak of random samples of fixed size r, the adjective random is used to imply that all possible samples have the same probability $1/n^r$ in sampling with replacement and $1/(n)_r$ in sampling without replacement.

3 Examples.

We consider random samples of size r with replacement taken from a population of the n elements $a_1, a_2, \ldots, a_n$. We are interested in the event A that in such a sample $(a_{j_1}, a_{j_2}, \ldots, a_{j_r})$ no element appears twice, that is, that our sample could have been obtained also by sampling without replacement. The last theorem shows that there exist n^r different samples in all, of which $(n)_r$ samples satisfy the stipulated condition. Assuming that all arrangements have equal probability $1/n^r$, we conclude that the probability of no repetition in our sample is

$$p = \frac{(n)_r}{n^r} = \frac{n(n-1)\cdots(n-r_1)}{n^r}$$

Example 3.1. **Random sampling numbers.** Let the population consist of the ten digits $0, 1, 2, \ldots, 9$. Every succession of five digits represents a sample of size $r = 5$ and we assume that each such arrangement has a probability 10^{-5}. By the above result, the probability that five consecutive random digits are all different is $p = (10)_5 \cdot 10^-5 = 0.3024$.

Example 3.2. If n balls are randomly placed into n cells, the probability that each cell will be occupied is $n!/n^n$. This probability is surprisingly small: for $n = 7$, it is only 0.00612. This means that if in a city, seven accidents occur each week, then (assuming that all possible distributions are equally likely, practically all weeks will contain days with two or more accidents, and on the average only one week out of 165 will show a uniform distribution of one accident per day. This example shows an unexpected characteristic of pure randomness. For $n = 6$, the probability $n!n^{-n} = 0.01543$. This shows how extremely improbably is that in six throws with a fair die, all faces turn up.

Example 3.3. **Birthdays.** The birthdays of r people form a sample of size r from the population of all days in the year. The years are not of equal length and we know that the birth rates are not quite constant throughout the year. However, in a first approximation, we may take a random selection of people as equivalent to a random selection of birthdays and consider the year as consisting of 365 days.

With these conventions, we can interpret equation (3.1) to the effect that that probability, p, that all r birthdays are different is

$$\begin{aligned} p &= \frac{(365)_r}{365^r} \\ &= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{r-1}{365}\right) \end{aligned}$$

Again the numerical consequences are astounding. Thus, for $r = 23$ people have $p < \frac{1}{2}$, that is, for 23 people, the probability that at least two people have common birthday exceeds $\frac{1}{2}$.

4 Subpopulations and Partitions.

As before, we use the term population of size n to denote an aggregate of n elements without regard to their order. Two populations are considered different if one contains an element not contained in the other. Choosing