7 Sampling

When all elements of a finite sample space are equally likely calculating probability often boils down to counting the number of ways of making some selection. Specifically, we are often interested in finding how many ways there are of choosing r elements from an n element set. This is called *sampling from an n element set*. The answer to this question depends on exactly what we mean by selection: is the order important and is repetition allowed. We assume throughout this section that X is a set with n elements.

Ordered selection with repetition allowed

If we make an ordered selection of r things from X and allow elements to be repeated) then the sample space is

$$S = X^r = \{(x_1, x_2, \ldots, x_r) : x_i \in X\},$$

and if $|X| = n$ we have

$$|S| = n^r.$$

This follows since we have n choices for x_1, for each of these choices there are n choices for x_2, for each of these there are n choices for x_3 and so on. For example, looking back to exercise sheet 1 question 1, we were sampling twice from a 6 element set with replacement and we had $S = \{1, 2, 3, 4, 5, 6\}^2$ and $|S| = 36$.

Ordered selection without repetition

If we make an ordered selection of r things from X and we do not allow elements to be repeated then the sample space is

$$S = \{(x_1, x_2, \ldots, x_r) : x_i \in X \text{ with } x_i \neq x_j \text{ for all } i \neq j\}.$$

To find the cardinality of this notice that there are n choices for x_1, for each of these choices there are $n-1$ choices for x_2, for each of these there are $n-2$ choices for x_3 and so on. Hence,

$$|S| = n(n-1)(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

(notice that there are r terms in this product).

An important special case of this is that there are $n!$ ways of making an ordered selection of n things from an n element set – that is there are $n!$ different orderings of an n element set.

Unordered selection without repetition

If we make an unordered selection of r things from X without repetition then the sample space is just the set of all subsets of S of cardinality r i.e.

$$S = \{Y \subseteq X : |Y| = r\}.$$

An *ordered* selection can be obtained by taking an element Y of this sample space and putting its elements in order. Since $|Y| = r$, the elements of Y can be ordered in $r!$ ways and so (using the formula for ordered selections with repetition) we have

$$|S| = \frac{n!}{(n-r)!\,r!}\,.$$

This expression is usually written as $\binom{n}{r}$ (read as "n choose r") and is called a *binomial coefficient*. By convention $\binom{n}{r} = 0$ when $r > n$.

Summary

We have shown the following

Theorem 7.1. *Let X be a set with n elements. Then:*
(a) the number of ways of making an ordered selection of r elements from X with repetitions allowed is n^r.
(b) the number of ways of making an ordered selection of r elements from X with no repetitions allowed is $n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}$.
(c) the number of ways of making an unordered selection of r elements from X with no repetitions allowed is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.

It is important when answering questions involving sampling that you read the question carefully and decide what sort of sampling is involved. There are many examples in your notes and on exercise sheets. Sometimes more than one kind of sampling can be used but you must be consistent.

The Binomial Theorem

The following is an application of counting selections which will crop up later in the course (and in other courses).

Theorem 7.2 (The Binomial Theorem)**.** $(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$

Proof. We have

$$\begin{aligned}(x+y)^n &= \underbrace{(x+y) \times (x+y) \times \cdots \times (x+y)}_{n \text{ terms}} \\ &= a_0 y^n + a_1 x y^{n-1} + a_2 x^2 y^{n-2} + \cdots + a_n x^n\end{aligned}$$

where a_r is the number of ways of making an unordered selection of r elements from an n element set without repetition (we are selecting r terms in the product from which to take an x rather than a y). So $a_r = \binom{n}{r}$ and the result follows. $\square$

Remark. *The binomial theorem can also be proved easily by induction on n.*