7.2 Introduction

In Chapter 6, we learned how to manipulate probabilities and to obtain marginal probability distributions from joint probability distributions. But what we do not yet know is how to obtain a probability mass function, either joint or marginal, in the first place; that is the subject of this chapter and its extension to continuous variables in the next. All we have been able to do so far is to assign probabilities on the "equally likely principle." But as we recall, we do this because we don't know what the true probability function really is. Our next step is to discover how different experiments or different surveys lead to different types of probability mass function.

The major lesson to be learned in this chapter is that different hypothetical experiments generate different types of probability distribution. Once the nature of the hypothetical experiment has been determined, that specification of the experiment will also specify a particular corresponding type of probability distribution. Consequently, we now realize that probability distributions do not spring out of the thin air, or from the fervid imaginations of statistics professors, but arise naturally from the specification of an experiment.

Ultimately, our objective is to come back to real data and the outcomes of actual experiments or surveys. At that time, we will benefit from our recognition that there is a strong link between experiment and the form of a probability distribution. And our first task will be to discover what probability distribution applies to the situation in hand.

In this chapter, we will discuss only a few probability distribution functions but enough to show clearly the link between hypothetical experiment and corresponding probability function. We will also try to demonstrate how a probability function will change as we change the experiment. Indeed, this dependence of the specific form of the distribution on the precise formulation of the experiment will be emphasized. Before we can begin we will have to learn some new notation and some new mathematical tools. This is our next topic.

7.3 Combinations and Permutations

Suppose that we have n distinct objects, $A_1, A_2, A_3, \ldots, A_n$. The objects might be different fruits, such as an apple, a pear, an orange, a banana, and so on. In how many distinct ways can we rearrange our n distinct objects, say our fruit, in a line? One arrangement is

Orange Banana Pear Apple

An alternative way of saying this is, "In how many distinct ways can we permute the order of our n distinct objects?" Suppose that we have only four fruits: an apple, a pear, an orange, and a banana. We also have four positions to fill:

________ ________ ________ ________

We have four choices for the first position. We might pick a banana, for example. But for each choice for the first position, we will have three choices for the second position. If our first choice was the banana, then for our second choice we have only the pear, orange, and an apple left. After that choice we have only two choices left for the next and third slot; and for the last slot, we have only one choice, in fact no choice at all.

We can generalize our discovery. If we have n distinct objects, then we have

n ways of picking the first in line
$(n - 1)$ ways of picking the second in line
$(n - 2)$ ways of picking the third in line
...
1 way of picking the last object

With each of the n ways of picking the first object, we have $(n - 1)$ ways of picking the second object, and with each of the $(n - 1)$ ways of picking the second object, we have $(n - 2)$ ways of picking the second object, and so on. Each possible choice for rearranging the distinct elements is called a **permutation**.

We conclude that the total number of permutations of n distinct objects is

$$n(n-1)(n-2)(n-3)\dots 1$$

The terms are multiplied because for each choice at every position, you can have any one of the remaining choices. This special product is called "n factorial" and is written "$n!$"; that is

$$n! = n(n-1)(n-2)(n-3)\dots 1$$

It is the simple product of all the numbers from 1 to n; for n distinct elements we have a total of $n!$ permutations.

In how many ways can we permute our four fruits? Because n takes the value "four," the answer is 4!, which equals 24.

Let's look at an even easier example but examine the alternative permutations specifically. Try the three letters A, B, and C. Their permutations are

$A\ B\ C$	$A\ C\ B$
$B\ A\ C$	$B\ C\ A$
$C\ A\ B$	$C\ B\ A$

With this example check that for each of the first three choices, you have two choices left; so that the number of permutations is 3 times 2. You might not relish permuting four different types of fruit, but if you relabel them as the letters A to D as we did here, you can permute them for yourself.

All this is easy enough, but we will also want to know how to figure out how many different ways we can select r objects taken from n distinct objects. This question would arise, for example, if we wanted to know how many ways we could select two fruits at a time from the set of four. What this is really saying is that, having selected our two objects, we do not care about the order of the remaining objects. How do we do this? We have two slots to fill, as shown, four choices for the first and three choices for the second slot.

________ ________

In general, if we have n objects and want to select r of them, then we have n choices for the first position, $(n-1)$ choices for the second position, $(n-2)$ choices for the third position, and so on:

$$n(n-1)(n-2)(n-3)\cdots(n-r+1)$$

Notice that as there are r terms, the index of the last term in the expansion is $(n-r+1)$; this is because the first term in the expansion is $(n-0)$.

Alternatively, if we do not care about the order of the remaining $(n-r)$ objects, even if we actually have them, then the number of distinct ways of selecting r objects out of n distinct objects is still

$$n(n-1)(n-2)(n-3)\cdots(n-r+1)$$

Returning to our four fruits, for example, if we want to select only two fruits, then $r = 2$, and $n = 4$; so the number of distinct choices is 4 times 3 = 12.

It is instructive and useful to reexpress the problem in a slightly different way. If we are not interested in the order, or permutations, of the remaining $(n-r)$ objects, and given that we multiply our options at each position, then we should divide out those choices that we do not make. Our solution for the number of ways of selecting r objects out of n objects is by this reasoning:

$$\begin{aligned} n! &= \text{total number of permutations of } n \text{ objects} \\ (n-r)! &= \text{number of permutations of } (n-r) \text{ objects} \end{aligned}$$

so the number of ways of selecting r objects and ignoring all permutations of the remaining $(n-r)$ objects not chosen is to divide out the unwanted permutations:

$$\begin{aligned} \text{No. of perms.} &= \frac{1 \times 2 \times 3 \times \cdots \times (n-r) \times (n-r+1) \times \cdots \times (n-1) \times n}{1 \times 2 \times 3 \times \cdots \times (n-r)} \\ &= (n-r+1) \times (n-r+2) \times \cdots \times (n-1) \times n \\ \frac{n!}{(n-r)!} &= \frac{1 \times 2 \times 3 \times \cdots \times (n-r) \times (n-r+1) \times \cdots \times (n-1) \times n}{1 \times 2 \times 3 \times \cdots \times (n-r)} \\ &= n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) \end{aligned}$$

Our answer is found by dividing out the unwanted permutations; the number to divide by is $(n-r)!$ for the $(n-r)!$ permutations of the $(n-r)$ remaining objects. The total number of objects is n so that the total number of permutations to be divided is $n!$. This is a good strategy to follow in general; figure out how many objects have permutations in which you are not interested and divide out by the appropriate factorial.

To make this idea work in full generality, we need a convention, an agreement, that

$$0! = 1$$

Let us return to our example of the four letters, A, B, C, and D. In how many ways can we choose two letters out of the four? Our expression yields

$$\frac{n!}{(n-r)!} = \frac{4!}{(4-2)!}$$
$$= \frac{4!}{(2)!}$$
$$= \frac{(1 \times 2 \times 3 \times 4)}{(1 \times 2)}$$
$$= 3 \times 4 = 12$$

by dividing out the common products, 1×2, in both the numerator and the denominator. To see this result clearly, let us write out the permutations:

AB CD AC AD BC BD
BA DC CA DA CB DB

This list shows the permutations in which we are interested. List the permutations in which we are not interested, the ones left out, and count them. You should have

$$(4! - 12) = 24 - 12 = 12$$

Yet another useful way of viewing this result is to answer the related question. How many distinct permutations are there for n objects of which $(n - r)$ are identical but the remaining r objects are distinct? Mathematically, this question is just like the previous question. You have n objects of which r are distinct and can be permuted, but the remaining $(n - r)$ objects are either the same or you are not interested in their permutations. In both cases, the total number of possible permutations with n objects is $n!$, but you want to divide out the $(n - r)!$ permutations that you cannot, or do not want to, distinguish. Let us consider a sample example of this alternative interpretation.

We have five objects, *A A A B C*, of which three are the same, so $n = 5$; $r = 2$; $(n - r) = 3$:

$$\frac{n!}{(n-r)!} = \frac{5!}{3!} = \frac{1 \times 2 \times 3 \times 4 \times 5}{1 \times 2 \times 3} = 4 \times 5 = 20$$

Here is the list:

A A A B C	*A A A C B*	*A A B C A*	*A A C B A*	*A B C A A*
A C B A A	*A A B A C*	*A A C A B*	*B C A A A*	*C B A A A*
B A A A C	*B A A C A*	*B A C A A*	*C A A A B*	*C A A B A*
C A B A A	*A B A A C*	*A B A C A*	*A C A A B*	*A C A B A*

Alternatively, if we have n objects, of which r are the same, then the number of distinct permutations is given by

$$\frac{n!}{r!}$$

But if we can calculate the number of permutations when either r or $(n - r)$ of them are the same, could we not calculate the number of distinct choices when there are both r and $(n - r)$ the same? We have n objects as before, but now we have r of them the same and $(n - r)$ of them the same but different from each other. Suppose that we have 4 *A*s and 3 *B*s. Here n is 7, r is 4, and $(n - r)$ is 3. Our expression tells us that out of a total of 7! permutations, we must divide out both the 4! and 3! permutations that we cannot distinguish. The expression is

$$\begin{aligned}\frac{n!}{r!(n-r)!} &= \frac{7!}{3!4!} \\ &= \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7}{(1 \times 2 \times 3) \times (1 \times 2 \times 3 \times 4)} \\ &= 5 \times 7 = 35\end{aligned}$$

The result of 35 you might find a little surprising with only two different letters to play with, but there are seven positions to fill!

This particular expression is used so often that it has a special name and a special symbol. The special name is **combination.** One asks how many combinations of n objects can be obtained if taken r at a time. Notice that with a combination we are not interested in the order of either the r items selected or the $(n - r)$ items not selected. If we know the total number of permutations for n objects, we can obtain the number of combinations by dividing out the unwanted permutations generated by the order within each selection and the order within each set not selected. The special symbol is shown in Equation 7.1:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{7.1}$$

Notice that the roles of r and $(n - r)$ can be interchanged with no effect on the value taken by the expression; that is

$$\binom{n}{n-r} = \frac{n!}{(n-r)!r!} = \binom{n}{r} \tag{7.2}$$

You might want to imagine that combinations can be obtained by placing the n objects into two boxes, one containing r objects, the other containing $(n - r)$ objects. The order of the objects within each box is unimportant.

The generic question with combinations is, "In how many ways can I separate n objects into two categories?" So we might consider how many committees of 5 can we select from 15 people; how many flower arrangements we can obtain from ten different flowers taken three at a time; how many ways there are to selecting three colors from six colors when producing maps; or how many car options there are if a consumer can select any 6 of them and there are 20 to choose from; and so on.