

PROBABILITY DISTRIBUTIONS: DISCRETE AND CONTINUOUS

Univariate Probability Distributions. Let S be a sample space with a probability measure P defined over it, and let x be a real scalar-valued set function defined over S , which is described as a random variable.

If x assumes only a finite number of values in the interval $[a, b]$, then it is said to be discrete in that interval. If x assumes an infinity of values between any two points in the interval, then it is said to be everywhere continuous in that interval. Typically, we assume that x is discrete or continuous over its entire domain, and we describe it as discrete or continuous without qualification.

If $x \in \{x_1, x_2, \dots\}$, is discrete, then a function $f(x_i)$ giving the probability that $x = x_i$ is called a *probability mass function*. Such a function must have the properties that

$$f(x_i) \geq 0, \quad \text{for all } i, \quad \text{and} \quad \sum_i f(x_i) = 1.$$

Example. Consider $x \in \{0, 1, 2, 3, \dots\}$ with $f(x) = (1/2)^{x+1}$. It is certainly true that $f(x_i) \geq 0$ for all i . Also,

$$\sum_i f(x_i) = \left\{ \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots \right\} = 1.$$

To see this, we may recall that

$$\frac{1}{1-\theta} = \{1 + \theta + \theta^2 + \theta^3 + \dots\},$$

whence

$$\frac{\theta}{1-\theta} = \{\theta + \theta^2 + \theta^3 + \dots\}.$$

Setting $\theta = 1/2$ in the expression above gives $\theta/(1-\theta) = \frac{1/2}{(1-1/2)} = 1$, which is the result that we are seeking.

If x is continuous, then a *probability density function* (p.d.f.) $f(x)$ may be defined such that the probability of the event $a < x \leq b$ is given by

$$P(a < x \leq b) = \int_a^b f(x)dx.$$

Notice that, when $b = a$, there is

$$P(x = a) = \int_a^a f(x)dx = 0.$$

That is to say, the integral of the continuous function $f(x)$ at a point is zero. This corresponds to the notion that the probability that the continuous random variable

x assumes a particular value amongst the non denumerable (i.e. uncountable) infinity of values within its domain must be zero. Nevertheless, the value of $f(x)$ at a point continues to be of interest; and we describe this as a probability measure as opposed to a probability.

Notice also that it makes no difference whether we include or exclude the endpoints of the interval $[a, b] = \{a \leq x \leq b\}$. The reason that we have chosen to use the half-open half-closed interval $(a, b] = \{a < x \leq b\}$ is that this accords best with the definition of a definite integral over $[a, b]$, which is obtained by subtracting the integral over $(-\infty, a]$ from the integral over $(-\infty, b]$.

Example. Consider the exponential function

$$f(x) = \frac{1}{a}e^{-x/a}$$

defined over the interval $[0, \infty)$ and with $\alpha > 0$. There is $f(x) > 0$ and

$$\int_0^{\infty} \frac{1}{a}e^{-x/a} dx = \left[-e^{-x/a}\right]_0^{\infty} = 1.$$

Therefore, this function constitutes a valid p.d.f. The exponential distribution provides a model for the lifespan of an electronic component, such as fuse, for which the probability of failing in the ensuing period is liable to be independent of how long it has survived already.

Corresponding to any p.d.f $f(x)$, there is a cumulative distribution function, denoted by $F(x)$, which, for any value x^* , gives the probability of the event $x \leq x^*$. Thus, if $f(x)$ is the p.d.f. of x , which we denote by writing $x \sim f(x)$, then

$$F(x^*) = \int_{-\infty}^{x^*} f(x) dx = P(-\infty < x \leq x^*).$$

In the case where x is a discrete random variable with a probability mass function $f(x)$, also denoted by $x \sim f(x)$, there is

$$F(x^*) = \sum_{x \leq x^*} f(x) = P(-\infty < x \leq x^*).$$

Pascal's Triangle and the Binomial Expansion. The binomial distribution is one of the fundamental distributions of mathematical statistics. To derive the binomial probability mass function from first principles, it is necessary to invoke some fundamental principles of combinatorial calculus. We shall develop the necessary results at some length in the ensuing sections before deriving the mass function.

Consider the following binomial expansions:

$$\begin{aligned} (p + q)^0 &= 1, \\ (p + q)^1 &= p + q, \\ (p + q)^2 &= p^2 + 2pq + q^2, \end{aligned}$$

$$\begin{aligned}
(p+q)^3 &= p^3 + 3p^2q + 3pq^2 + q^3, \\
(p+q)^4 &= p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4, \\
(p+q)^5 &= p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5.
\end{aligned}$$

The generic expansion is in the form of

$$\begin{aligned}
(p+q)^n &= p^n + np^{n-1}q + \frac{n(n-1)}{2}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{3!}p^{n-3}q^3 + \\
&\quad \dots + \frac{n(n-1)\dots(n-r+1)}{r!}p^{n-r}q^r + \dots \\
&\quad + \frac{n(n-1)(n-2)}{3!}p^3q^{n-3} + \frac{n(n-1)}{2}p^2q^{n-2} + npq^{n-1} + p^n.
\end{aligned}$$

In a tidier notation, this becomes

$$(p+q)^n = \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x q^{n-x}.$$

We can find the coefficient of the binomial expansions of successive degrees by the simple device known as Pascal's triangle:

$$\begin{array}{cccccc}
& & & & & & 1 \\
& & & & & & & 1 \\
& & & & & & 1 & & 1 \\
& & & & & & 1 & 2 & 1 \\
& & & & & 1 & 3 & 3 & 1 \\
& & & & 1 & 4 & 6 & 4 & 1 \\
& & 1 & 5 & 10 & 10 & 5 & 1
\end{array}$$

The numbers in each row but the first are obtained by adding two adjacent numbers in the row above. The rule is true even for the units that border the triangle if we suppose that there are some invisible zeros extending indefinitely on either side of each row.

Instead of relying merely upon observation to establish the formula for the binomial expansion, we should prefer to derive the formula by algebraic methods. Before we do so, we must reaffirm some notions concerning permutations and combinations that are essential to a proper derivation.

Permutations. Let us consider a set of three letters $\{a, b, c\}$ and let us find the number of ways in which they can be arranged in a distinct order. We may pick any one of the three to put in the first position. Either of the two remaining letters may be placed in the second position. The third position must be filled by the unused letter. With three ways of filling the first place, two of filling the second

and only one way of filling the third, there are altogether $3 \times 2 \times 1 = 6$ different arrangements. These arrangements or permutations are

$$abc, \quad cab, \quad bca, \quad cba, \quad bac \quad acb.$$

Now let us consider an unordered set of n objects denoted by

$$\{x_i; i = 1, \dots, n\},$$

and let us ascertain how many different permutations arise in this case. The answer can be found through a litany of questions and answers which we may denote by $[(Q_i, A_i); i = 1, \dots, n]$:

Q_1 :	In how many ways can the first place be filled?	A_1 : n
Q_2 :	In how many ways can the second place be filled?	A_2 : $n - 1$ ways,
Q_3 :	In how many ways can the third place be filled?	A_3 : $n - 2$ ways,
\vdots		
Q_r :	In how many ways can the r th place be filled?	A_r : $n - r + 1$ ways,
\vdots		
Q_n :	In how many ways can the n th place be filled?	A_n : 1 way.

If the concern is to distinguish all possible orderings, then we can recognise

$$n(n - 1)(n - 2) \cdots 3 \cdot 2 \cdot 1 = n!$$

different permutations of the objects. We call this number n -factorial, which is written as $n!$, and it represents the number of permutations of n objects taken n at a time. We also denote this by

$${}^n P_n = n!.$$

We shall use the same question-and-answer approach in deriving several other important results concerning permutations. First we shall ask

Q: How many ordered sets can we recognise if r of the n objects are so alike as to be indistinguishable?

A reasoned answer is as follows:

A: Within any permutation there are r objects that are indistinguishable. We can permute the r objects amongst themselves without noticing any differences. Thus the suggestion that there might be $n!$ permutations would overestimate the number of distinguishable permutations by a factor of $r!$; and, therefore, there are only

$$\frac{n!}{r!} \quad \text{recognizably distinct permutations.}$$

Q: How many distinct permutations can we recognise if the n objects are divided into two sets of r and $n - r$ objects—eg. red billiard balls and white billiard balls—where two objects in the same set are indistinguishable?

A: By extending the previous argument, we should find that the answer is

$$\frac{n!}{(n-r)!r!}.$$

Q: How many ways can we construct a permutation of r objects selected from a set of n objects?

A: There are two ways of reaching the answer to this question. The first is to use the litany of questions and answers which enabled us to discover the total number of permutations of n objects. This time, however, we proceed no further than the question Q_r , for the reason that there are no more than r places to fill. The answer we seek is the number of way of filling the first r places. The second way is to consider the number of distinct permutation of n objects when $n-r$ of them are indistinguishable. We fail to distinguish amongst these objects because they all share the same characteristic which is that they have been omitted from the selection. Either way, we conclude that the number is

$$\begin{aligned} {}^n P_r &= n(n-1)(n-2)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

Combinations. Combinations are selections of objects in which no attention is paid to the ordering. The essential result is found in answer to just one question:

Q: How many ways can we construct an unordered set of r objects selected from amongst n objects?

A: Consider the total number of permutations of r objects selected from amongst n . This is ${}^n P_r$. But each of the permutations distinguishes the order of the r objects it comprises. There are $r!$ different orderings or permutations of r objects; and, to find the number of different selections or combinations when no attention is paid to the ordering, we must deflate the number ${}^n P_r$ by a factor of $r!$. Thus the total number of combinations is

$$\begin{aligned} {}^n C_r &= \frac{1}{r!} {}^n P_r = \frac{n!}{(n-r)!r!} \\ &= {}^n C_{n-r}. \end{aligned}$$

Notice that we have already derived precisely this number in answer to a seemingly different question concerning the number of recognizably distinct permutations of a set of n objects of which r were in one category and $n-r$ in another. In the present case, there are also two such categories: the category of those objects which are included in the selection and the category of those which are excluded from it.

Example. Given a set of n objects, we may define a so-called power set, which is the set of all sets derived by making selections or combinations of these objects. It is straightforward to deduce that there are exactly 2^n objects in the power set.

This is demonstrated by setting $p = q = 1$ in the binomial expansion above to reach the conclusion that

$$2^n = \sum_x \binom{n}{x} = \sum_x \frac{n!}{x!(n-x)!}.$$

Each element of this sum is the number of ways of selecting x objects from amongst n , and the sum is for all values of $x = 0, 1, \dots, n$.

The Binomial Theorem. Now we are in a position to derive our conclusion regarding the binomial theorem without, this time, having recourse to empirical induction. Our object is to determine the coefficient associated with the generic term $p^x q^{n-x}$ in the expansion of

$$(p + q)^n = (p + q)(p + q) \cdots (p + q),$$

where the RHS displays the n factors that are to be multiplied together. The coefficients of the various elements of the expansion are as follows:

- p^n The coefficient is unity, since there is only one way of choosing n of the p 's from amongst the n factors.
- $p^{n-1}q$ This term is the product of q selected from one of the factors and $n-1$ p 's provided by the remaining factors. There are $n = {}^nC_1$ ways of selecting the single q .
- $p^{n-2}q^2$ The coefficient associated with this term is the number of ways of selecting two q 's from n factors, which is ${}^nC_2 = n(n-1)/2$.
- \vdots
- $p^{n-r}q^r$ The coefficient associated with this terms is the number of ways of selecting r q 's from n factors, which is ${}^nC_r = n!/\{(n-r)!r!\}$.

From such reasoning, it follows that

$$\begin{aligned} (p + q)^n &= p^n + {}^nC_1 p^{n-1}q + {}^nC_2 p^{n-2}q^2 + \\ &\quad \cdots + {}^nC_r p^{n-r}q^r + \cdots \\ &\quad + {}^nC_{n-2} p^2 q^{n-2} + {}^nC_{n-1} p q^{n-1} + q^n \\ &= \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x q^{n-x}. \end{aligned}$$

The Binomial Probability Distribution. We wish to find, for example, the number of ways of getting a total of x heads in n tosses of a coin. First we consider a single toss of the coin. Let us take the i th toss, and let us denote the outcome by $x_i = 1$ if it is heads and by $x_i = 0$ if it is tails. Heads might be described as a success, whence the probability of a success will be $P(x_i = 1) = p$. Tails might be described as a failure and the corresponding probability of this event is $P(x_i = 0) = 1 - p$. For a fair coin, we should have $p = 1 - p = \frac{1}{2}$, of course.

We are now able to define a probability function for the outcome of the i th trial. This is

$$f(x_i) = p^{x_i}(1-p)^{1-x_i} \quad \text{with } x_i \in \{0, 1\}.$$

The experiment of tossing a coin once is called a Bernoulli trial, in common with any other experiment with a random dichotomous outcome. The corresponding probability function is called a point binomial.

If the coin is tossed n times, then the probability of any particular sequence of heads and tails, denoted by (x_1, x_2, \dots, x_n) is given by

$$\begin{aligned} P(x_1, x_2, \dots, x_n) &= f(x_1)f(x_2) \cdots f(x_n) \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\ &= p^x(1-p)^{n-x}, \end{aligned}$$

where we have defined $x = \sum x_i$. This result follows from the independence of the Bernoulli trials, whereby $P(x_i, x_j) = P(x_i)P(x_j)$ is the probability of the occurrence of x_i and x_j together in the sequence.

Altogether, there are

$$\binom{n}{x} = \frac{n!}{(n-x)!x!} = nC_x$$

different sequences (x_1, x_2, \dots, x_n) which contain x heads; so the probability of the event of x heads in n tosses is given by

$$b(x; n, p) = \frac{n!}{(n-x)!x!} p^x q^{n-x}.$$

This is called the binomial probability mass function.