Probability distributions

(Notes are heavily adapted from Harnett, Ch. 3; Hayes, sections 2.14-2.19; see also Hayes, Appendix B.)

I. Random variables (in general)

A. So far we have focused on single events, or with a combination of events in an experiment. Now we shall talk about the probability of all events in an experiment.

B. Imagine that each and every possible elementary event in the sample space S is assigned a number. That is, various elementary events are paired with various values of a variable.

- an elementary event might be a person, with some height in inches
- the elementary event may be the result of tossing a pair of dice, with the assigned number being the total of the spots that came up
- the elementary event may be a rat, with the number standing for the trials taken to learn a maze.

Each and every elementary event is thought of as getting one and only one such number.

C. Note that the elementary events themselves, and the values of the random variables associated with them, are <u>not</u> the same thing.

- For example, you might have a sample space which consists of all American males aged 21 and over - each such male is an elementary event in this sample space. Now we can associate with each elementary event a real value, such as the income of the man during the current calendar year. The values that the random variable X can thus assume are the various income values associated with the men. The particular value x occurs when a man is chosen who has income x.

D. <u>Random variable</u> - Let X represent a function that associates a real number with each and every elementary event in some sample space S. Then X is called a <u>random variable</u> on the sample space S. Chance variable and stochastic variable are alternative terms. Harnett uses the alternative but equivalent definition that a Random Variable is a well-defined rule for assigning a numerical value to every possible outcome of an experiment.

E. EXAMPLES:

- Coin flip. X = 1 if heads, 0 otherwise.
- Height. X = height, measured to the nearest inch.

F. <u>Notation</u>. Typically, capital letters, such as X, Y, and Z, are used to denote random variables, and lowercase letters, such as x, y, z and a, b, c are used to denote *particular values* that the random variable can take on. Thus, the expression P(X = x) symbolizes the probability that the random variable X takes on the particular value x. Often, this is written simply as P(x). Likewise, $P(X \leq x)$ = probability that the random variable X is less than or equal to the specific value x; $P(a \leq X \leq b)$ = probability that X lies between values a and b. Harnett, on the other hand, likes to use bold-face italic for rvs, and hence in his notation P(***x*** = x) symbolizes the probability that the random variable ***x*** takes on the particular value x.

II. Discrete random variables

A. In a great many situations, only a limited set of numbers can occur as values of a random variable. Quite often, the set of numbers that can occur is relatively small, or at least finite in extent.

For example, suppose I randomly draw a page from the statistics book and note the page number. In this instance, the values of the random variable are all of the different page numbers that might occur.

Some random variables can assume what is called a “countably infinite” set of values. One example of a countably infinite set would be the ordinary counting numbers themselves, where the count goes on without end. A simple experiment in which one counts the number of trials until an event occurs would give a random variable taking on these counting values, e.g. flipping a coin until a heads comes up.

B. Discrete random variable - in either of these situations, the random variable is said to be discrete. If a random variable X can assume only a particular finite or countably infinite set of values, it is said to be a discrete random variable. Not all random variables are discrete, but a large number of random variables of practical and theoretical interest to us will have this property.

C. Continuous random variable. By way of contrast, consider something like height, which can take on an infinite, non-countable number of values (e.g. 6.0 feet, 6.01 feet, 6.013 feet, 6.2 feet, 6.204 feet, etc.). Variables such as height are continous.

To put it another way - discrete variables tend to be things you count, while continuous variables tend to be things you measure.

As we will see later, we can often treat variables as continuous even though they may be discrete and finite. For example, the number of unemployed workers in the U.S. is technically discrete and finite (though very large). But, statistically, it is easier to work with such a variable by treating it as continuous.

D. A Probability Distribution is a specification (in the form of a graph, a table or a function) of the probability associated with each value of a random variable.

E. Probability Mass Function = A probability distribution involving only discrete values of X. Graphically, this is illustrated by a graph in which the x axis has the different possible values of X, the Y axis has the different possible values of P(x).

Properties:

$0 \leq P(X = x) \leq 1$

$\Sigma\, P(X = x) = 1.$

F. Cumulative Distribution Function: The probability that a random variable X takes on a value less than or equal to some particular value *a* is often written as

$$F(a) = p(X \leq a) = \sum_{X \leq a} p(x) \text{ (for discrete variables)}$$

G. EXAMPLE – DISCRETE CASE. Probability calculations are often very simple when one is dealing with a discrete random variable where only a very few values can occur. See Hayes, pp. 95-96, for an example of an experiment involving rolling two dice. Here is another example.

Consider the simple experiment of tossing a coin three times. Let X = number of times the coin comes up heads. The 8 possible elementary events, and the corresponding values for X, are:

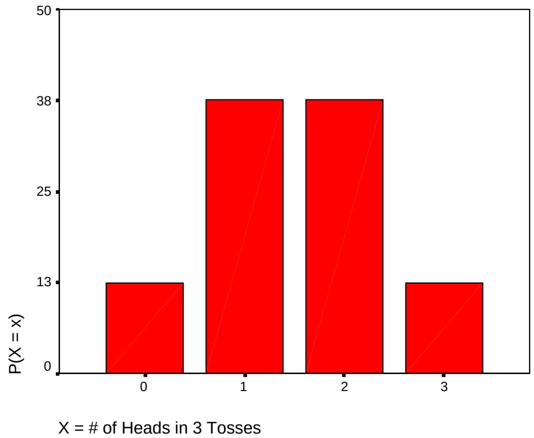
Elementary event	Value of X
TTT	0
TTH	1
THT	1
HTT	1
THH	2
HTH	2
HHT	2
HHH	3

Therefore, the probability distribution for the number of heads occurring in three coin tosses is:

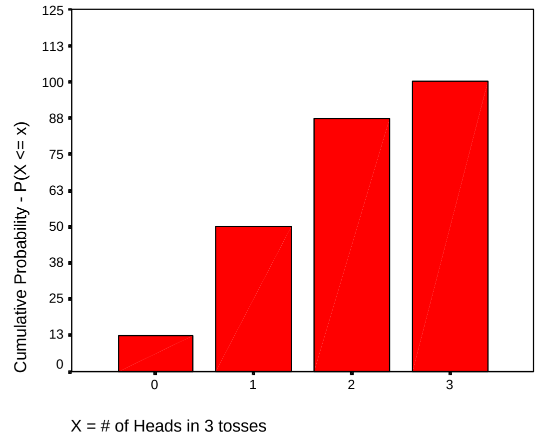
x	p(x)	F(x)
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

Graphically, we might depict this as

Probability Mass Function



Cumulative Distribution Function



From the above, any number of other questions can be answered. For example, $P(1 \le X \le 3)$ (i.e. the probability that you will get at least one head) = P(1) + P(2) + P(3) = 3/8 + 3/8 + 1/8 = 7/8. Or, if you prefer, you can use the complements rule and note that P(at least 1 head) = 1 - P(No heads) = 1 - 1/8 = 7/8.

G. <u>Function rules</u> - Sometimes it is easier to specify the distribution of a discrete random variable by its rule, rather than by a simple listing or graph. For example, suppose X = the value obtained by tossing a fair die. The function rule is

$$P(x) = \begin{cases} 1/6 & \text{if } x = 1,2,3,4,5,6 \\ 0 & \text{Otherwise} \end{cases}$$

A rule for our coin-tossing experiment would be:

$$P(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1,2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{Otherwise} \end{cases}$$

III. Continuous random variables.

A. Suppose that we could measure height to any degree of precision, regardless of how many decimal points that would require. Height would then be measured as a <u>continuous</u> variable. Continuous variables are treated somewhat differently than discrete variables. But fortunately, most probability theory is basically the same for discrete and continuous variables.

NOTE: A discussion of continuous variables requires we use some notation from calculus – but don't worry! You won't actually need to do any calculus yourself. As we'll see, most of the key calculations that require calculus have already been done for you and placed into tables.

B. Properties:

1. In general, for continuous random variables, the occurrence of any exact value of X may be regarded as having zero probability. For this reason, one does not usually discuss the probability per se for a value of a continuous random variable. Instead of the probability that X takes on some value a, we deal with the so-called probability density of X at a, symbolized by

$$f(a) = \text{probability density of } X \text{ at } a$$

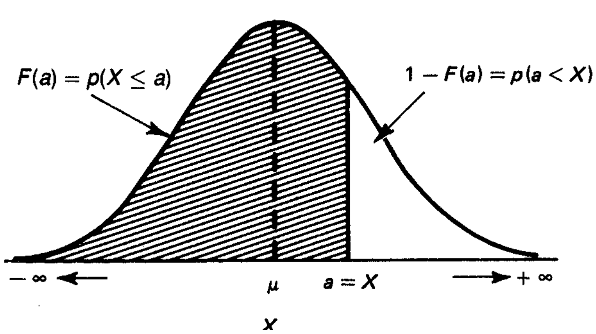
2. However, intervals of values can always be assigned probabilities. The probability of any continuous interval is given by

$$p(a \le X \le b) = \int_a^b f(x)\,dx = \text{Area under f(X) from a to b}$$

That is, the probability of an interval is the same as the area cut off by that interval under the curve for the probability densities, when the random variable is continuous and the total area is equal to 1.00.

C. For a continuous variable, the cumulative distribution function is written as

$$F(a) = p(X \le a) = \int_{-\infty}^{a} f(x)\,dx = \text{Area up to X = a}$$



The probability that a continuous random variable takes on any value between limits a and b can be found from

$$p(a \le X \le b) = F(b) - F(a)$$

This is seen easily if it is recalled that F(b) is the probability that X takes on value b or below, F(a) is the probability that X takes on value a or below; their difference must be the probability of a value between a and b.