Subject: **Introduction to Statistics**

BS Mathematics Morning/Evening program spring semester 2020

6

Probability

6.1 INTRODUCTION

The word *probability* has two basic meanings: (i) a quantitative measure of uncertainty and (ii) a measure of degree of belief in a particular statement or problem.

Probability and statistics are fundamentally interrelated. Probability is often called the vehicle of statistics. The area of inferential statistics in which we are mainly concerned with drawing inferences from experiments or situations involving an element of uncertainty, leans heavily upon probability theory. Uncertainty is also an inherent part of statistical inference as inferences are based on a sample, and a sample being a small part of the larger population, contains incomplete information. A similar type of uncertainty occurs when we toss a coin, draw a card or throw dice, etc. The uncertainty in all these cases is measured in terms of probability.

It is always clear what we mean when we make statements of the type that it is very likely to rain today or I have a fair chance of passing the annual examination or *A* will probably win a prize, etc. In each of these statements, the natural state of uncertainty is expressed, but on the basis of past evidence, we have some degree of personal belief in the truth of each statement.

The foundations of probability were laid by two French mathematicians of the seventeenth century–Blaise Pascal (1623-1662) and Pierre De Fermat (1601-1665)–in connection with gambling problems. Later on it was developed by Jakob Bernoulli (1654-1705), Abraham De Moivre (1667-1754) and Pierre Simon Laplace (1749-1827). The modern treatment of probability theory which consists of stating a few axioms and rules resulting from these axioms, was developed during the twenties and thirties of this century.

6.3 RANDOM EXPERIMENT

The term *experiment* means a planned activity or process whose results yield a set of data. A single performance of an experiment is called a *trial*. The result obtained from an experiment or a trial is called an *outcome*.

An experiment which produces different results even though it is repeated a large number of times under essentially similar conditions, is called a *random experiment*. The tossing of a fair coin, the throwing of a balanced die, drawing of a card from a well-shuffled deck of 52 playing cards, selecting a sample, etc. are examples of random experiments. A random experiment has three properties:

(i) The experiment can be repeated, practically or theoretically, any number of times,

(ii) The experiment always has two or more possible outcomes. An experiment that has only one possible outcome, is not a random experiment.

(iii) The outcome of each repetition is unpredictable, *i.e.* it has some degree of uncertainty.

It is to be remembered that an ordinary deck of playing cards contains 52 cards arranged in 4 suits of 13 each. The four suits are called *diamonds, hearts, clubs* and *spades*; the first two are red and the last two are black. The face values called *denominations,* of the 13 cards in each suit are ace, 2, 3, ..., 10, jack, queen and king. The term *honour card* refers to the denominations ace, 10, jack, queen and king. The *face cards* are king, queen and jack. These cards are used for various games such as whist, bridge, poker, etc.

6.3.1 Sample Space. A set consisting of all possible outcomes that can result from a random experiment (real or conceptual), is defined to be a *sample space* for the experiment and is denoted by the letter S. Each possible outcome is a member of the sample space and is called a *sample point* in that space. For instance, the experiment of tossing a coin results in either of the two possible outcomes: a head (H) or a tail (T); landing on its edge or rolling away is not considered. The sample space for this experiment may be expressed in set notation as $S = \{H, T\}$. The sample

6.6 CONDITIONAL PROBABILITY

The sample space for an experiment must often be changed when some additional information pertaining to the outcome of the experiment is received. The effect of such information is to *reduce* the sample space by excluding some outcomes as being impossible which before receiving the information were believed possible. The probabilities associated with such a *reduced* sample space are called *conditional probabilities*. The following example illustrates the concept of conditional probability.

Let us consider the die-throwing experiment with sample space $S = \{1, 2, 3, 4, 5, 6\}$. Suppose we wish to know the probability of the outcome that the *die shows 6*, say event A. If before seeing the outcome, we are told that the die shows an even number of dots, say event B, then the information that the *die shows an even number* excludes the outcomes 1, 3 and 5, and thereby *reduces* the original sample space to a sample space that consists of 3 outcomes 2, 4 and 6, *i.e.* the *reduced*

Example 6.33 If 60 percent of the voters in the City of Lahore prefer candidate X, what is the probability that in a sample of 12 voters exactly 7 will prefer X?

Here $n = 12, k = 7, p = 0.60$ and $q = 0.40$

Therefore $P(7 \text{ out of } 12 \text{ prefer } X) = \binom{12}{7}(0.60)^7 (0.40)^{12-7}$

$= (792)(0.02799)(0.01024) = 0.227$

Theorem 6.11 Bayes' Theorem. If the events $A_1, A_2, \ldots, A_k$ form a partition of a sample space S, that is, the events A_i are mutually exclusive and their union is S, and if B is any other event of S such that it can occur only if one of the A_i occurs, then for any i,

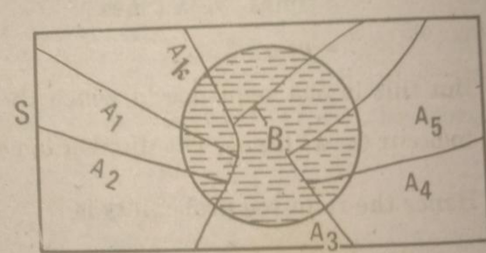
$$P(A_i/B) = \frac{P(A_i)\,P(B/A_i)}{\sum_{i=1}^{k} P(A_i)\,P(B/A_i)}, \quad \text{for } i = 1, 2, \ldots, k.$$

Proof: By the multiplicative law of probabilities, we have

$$P(B \cap A_i) = P(B)\,P(A_i/B)$$
$$= P(A_i)\,P(B/A_i).$$

Equating the equivalent relations of $P(B \cap A_i)$ and solving for $P(A_i/B)$, we get

$$P(A_i/B) = \frac{P(A_i)\,P(B/A_i)}{P(B)}$$



B is shaded

We may write the event B as $B = S \cap B$ (see the Venn diagram)

$= (A_1 \cup A_2 \cup \ldots \cup A_k) \cap B$

$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_k \cap B),$

where the $A_i \cap B$ are also mutually exclusive.

Therefore $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_k \cap B)$

Using the multiplicative law of probabilities, we may express each term $P(A_i \cap B)$ as $P(A_i)\,P(B/A_i)$. Then

$$P(B) = P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2) + \ldots + P(A_k)\,P(B/A_k)$$
$$= \sum_{i=1}^{k} P(A_i)\,P(B/A_i)$$

This result is generally known as the *theorem on total probability*. Replacing $P(B)$ by the total probability formula for the event B, we obtain *Bayes' formula* as

$$P(A_i/B) = \frac{P(A_i)\,P(B/A_i)}{\sum_{i=1}^{k} P(A_i)\,P(B/A_i)}$$

This result is known as Bayes' theorem after an English clergyman Thomas Bayes (1702-1761) who derived it and first used in a paper that was published posthumously in 1763. It should be noted that the original probabilities $P(A_i)$ are known as the *a priori* probabilities and the conditional probabilities $P(A_i/B)$ are called the *a posteriori* or *inverse* probabilities because probabilities are revised after some additional information has been obtained. Bayes' formula is also called the *formula for probabilities of hypotheses* on account of the reason that the events $A_1, A_2, ..., A_k$ may be thought of as hypotheses to account for occurrence of the event B.

Example 6.34 In a bolt factory, machines A, B and C manufacture 25, 35 and 40 percent of the total output, respectively. Of their outputs, 5, 4, and 2 percent, respectively, are defective bolts. A bolt is selected at random and found to be defective. What is the probability that the bolt came from machine A? B? C?

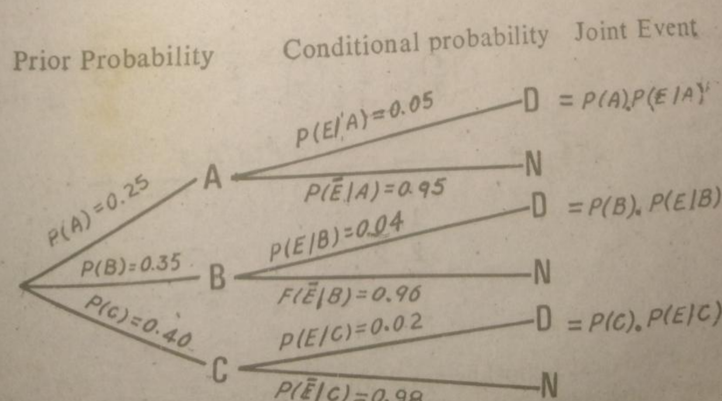
The *a priori* probabilities (before the information that the bolt is defective) are $P(A) = 0.25$, $P(B) = 0.35$, and $P(C) = 0.40$.

Let E represent the event that a bolt is defective (D).

Then the conditional probabilities are

$$P(E/A) = 0.05,\ P(E/B) = 0.04, \text{ and } P(E/C) = 0.02.$$

The outcomes with their respective probabilities may be shown by a tree diagram as below:



Now $P(A/E)$ is the *a posteriori* probability that the selected defective bolt came from machine A. Therefore, by Bayes' theorem, we get

$$P(A/E) = \frac{P(A).\,P(E/A)}{P(A).\,P(E/A) + P(B).\,P(E/B) + P(C).\,P(E/C)}$$

$$= \frac{(0.25)\,(0.05)}{(0.25)\,(0.05) + (0.35)\,(0.04) + (0.40)\,(0.02)}$$

$$= \frac{0.0125}{0.0345} = 0.362$$

Similarly, the posterior probabilities of machines B and C are

$$P(B/E) = 0.406, \text{ and } P(C/E) = 0.232$$

Example 6.35 An urn contains four balls which are known to be either (i) all white or (ii) two white and two black. A ball is drawn at random and is found to be white. What is the probability that all the balls are white? (P.U., B.A./B.Sc. (Hons.) Part III, 1966, 68)

Let A_1 be the hypothesis that all the balls are white and A_2 be the hypothesis that two are white and two black. Then the *a priori* probabilities must be

$P(A_1) = P(A_2) = \frac{1}{2}$, as the selection of a hypothesis is random.

Again let B be the event that the ball drawn is white. Then the conditional probabilities are

$$P(B/A_1) = \frac{{}^4C_1}{{}^4C_1} = 1 \text{ and } P(B/A_2) = \frac{{}^2C_1}{{}^4C_1} = \frac{1}{2}.$$

Therefore by Bayes' theorem, we get the *a posteriori* probabilities

$$P(A_1/B) = \frac{P(A_1)\,P(B/A_1)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2)}$$

$$= \frac{\left(\frac{1}{2}\right)(1)}{\left(\frac{1}{2}\right)(1) + \frac{1}{2}\left(\frac{1}{2}\right)} = \frac{1}{2} \times \frac{4}{3} = \frac{2}{3}; \text{ and}$$

$$P(A_2/B) = \frac{P(A_2)\,P(B/A_2)}{P(A_1)\,P(B/A_1) + P(A_2)\,P(B/A_2)}$$

$$= \frac{\frac{1}{2}.\frac{1}{2}}{\frac{1}{2}.1 + \frac{1}{2}.\frac{1}{2}} = \frac{1}{4} \times \frac{4}{3} = \frac{1}{3}.$$

Hence the first hypothesis, *i.e.* all the balls are white, is preferred as it has larger posterior probability.

	y_1	y_2	...	y_j	...	y_n	$P(X=x_i)$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	...	$f(x_1, y_j)$	...	$f(x_1, y_n)$	$g(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	...	$f(x_2, y_j)$	...	$f(x_2, y_n)$	$g(x_2)$
$\vdots$	$\vdots$					$\vdots$	$\vdots$
x_i	$f(x_i, y_1)$	$f(x_i, y_2)$	...	$f(x_i, y_j)$	...	$f(x_i, y_n)$	$g(x_i)$
$\vdots$	$\vdots$					$\vdots$	$\vdots$
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	...	$f(x_m, y_j)$	...	$f(x_m, y_n)$	$g(x_m)$
$P(Y=y_j)$	$h(y_1)$	$h(y_2)$	...	$h(y_j)$	...	$h(y_n)$	1

or be expressed by means of a formula for $f(x, y)$. The probabilites $f(x, y)$ can be obtained by substituting appropriate values of x and y in the table or formula.

A joint probability function has the following properties:

(i) $f(x_i, y_j) > 0$, for all (x_i, y_j), *i.e.* for $i = 1, 2, ..., m; j = 1, 2, ..., n.$

(ii) $\sum_i \sum_j f(x_i, y_j) = 1$

7.5.3 Marginal Probability Functions. From the joint probability function for (X, Y), we can obtain the individual probability function of X and Y. Such individual probability functions are called *marginal probability functions.*

Let $f(x, y)$ be the joint probability function of two discrete r.v.'s X and Y. Then the *marginal* probability function of X is defined as

$$g(x_i) = \sum_{j=1}^{n} f(x_i, y_j)$$

$$= f(x_i, y_1) + f(x_i, y_2) + ... + f(x_i, y_n) \text{ as } x_i \text{ must occur either with } y_1 \text{ or } y_2 \text{ or ... or } y_n.$$

$$= P(X=x_i);$$

that is, the individual probability function of X is found by adding over the rows of the two-way table.

Similarly, the *marginal* probability function for Y is obtained by adding over the column as

$$h(y_j) = \sum_{i=1}^{m} f(x_i, y_j) = P(Y=y_j)$$

(ii) $P(X + Y \leq 1)$;

(iii) the marginal p.d. $g(x)$ and $h(y)$;

(iv) the conditional p.d. $f(x \mid 1)$,

(v) $P(X=0 \mid Y=1)$, and

(vi) Are X and Y independent?

(i) The sample space S for this experiment contains $\binom{8}{2} = 28$ sample points. The possible values of X are 0, 1 and 2, and those for Y are 0, 1 and 2. The values that (X, Y) can take on are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2) and (2, 0). We desire to find $f(x, y)$ for each value (x, y).

Now $f(0, 0) = P(X=0 \text{ and } Y=0)$, where the event ($X=0$ and $Y=0$) represents that neither black nor red ball is selected, implying that the 2 selected are green balls. This event therefore contains $\binom{3}{0}\binom{2}{0}\binom{3}{2} = 3$ sample points, and

$$f(0, 0) = P(X=0 \text{ and } Y=0) = \frac{3}{28}.$$

Again $f(0, 1) = P(X=0 \text{ and } Y=1)$

$= P(\text{none is black, 1 is red and 1 is green})$

$$= \frac{\binom{3}{0}\binom{2}{1}\binom{3}{1}}{28} = \frac{6}{28}$$

Similarly, $f(1, 1) = P(X=1 \text{ and } Y=1)$

$= P(\text{1 is black, 1 is red and none is green})$

$$= \frac{\binom{3}{1}\binom{2}{1}\binom{3}{0}}{28} = \frac{6}{28}$$

Similar calculations give the probabilities of other values and the joint p.f. of X and Y is given as

(x, y)	(0, 0)	(0, 1)	(1, 0)	(1, 1)	(0, 2)	(2, 0)
$f(x, y)$	$\frac{3}{28}$	$\frac{6}{28}$	$\frac{9}{28}$	$\frac{6}{28}$	$\frac{1}{28}$	$\frac{3}{28}$

These probabilities can also be represented in another tabular form as follows:

Joint Probability Distribution

X \ Y	0	1	2	$P(X=x_i)$ $g(x)$
0	$\frac{3}{28}$	$\frac{6}{28}$	$\frac{1}{28}$	10/28
1	$\frac{9}{28}$	$\frac{6}{28}$	0	15/28
2	$\frac{3}{28}$	0	0	3/28
$P(Y=y_j)$ $h(y)$	15/28	12/28	1/28	1

Clearly, this joint p.d. of the two r.v.'s (X, Y) can be represented by the formula

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{28}, \qquad \begin{array}{l} x = 0, 1, 2 \\ y = 0, 1, 2 \\ 0 \le x + y \le 2. \end{array}$$

(ii) To compute $P(X + Y \le 1)$, we see that $x + y \le 1$ for the cells (0, 0), (0, 1) and (1, 0). Therefore

$$P(X + Y \le 1) = f(0, 0) + f(0, 1) + f(1, 0)$$

$$= \frac{3}{28} + \frac{6}{28} + \frac{9}{28} = \frac{18}{28} = \frac{9}{14}$$

(iii) The marginal p.d.'s are:

x	0	1	2
$g(x)$	10/28	15/28	3/28

y	0	1	2
$h(y)$	15/28	12/28	1/28

(iv) By definition the conditional p.d. $f(x \mid 1)$ is

$$f(x \mid 1) = P(X=x \mid Y=1)$$

$$= \frac{P(X=x \text{ and } Y=1)}{P(Y=1)} = \frac{f(x, 1)}{h(1)}$$

Now $$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{6}{28} + \frac{6}{28} + 0 = \frac{12}{28} = \frac{3}{7}$$

Therefore $$f(x \mid 1) = \frac{f(x, 1)}{h(1)} = \frac{7}{3} f(x, 1), \; x = 0, 1, 2$$

That is, $$f(0 \mid 1) = \frac{7}{3} f(0, 1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

$$f(1 \mid 1) = \frac{7}{3} f(1, 1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

$$f(2 \mid 1) = \frac{7}{3} f(2, 1) = \left(\frac{7}{3}\right)(0) = 0$$

Hence the conditional p.d. of X given that $Y=1$, is

x	0	1	2
$f(x \mid 1)$	1/2	1/2	0

(v) Finally, $P(X=0 \mid Y=1) = f(0 \mid 1) = \frac{1}{2}$

(vi) We find that $f(0, 1) = \frac{6}{28}$,

$$g(0) = \sum_{y=0}^{2} f(0, y) = \frac{3}{28} + \frac{6}{28} + \frac{1}{28} = \frac{10}{28}$$

$$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{6}{28} + \frac{6}{28} + 0 = \frac{12}{28}$$

Now $\quad \frac{6}{28} \neq \frac{10}{28} \times \frac{12}{28}$,

i.e. $\quad f(0, 1) \neq g(0)\, h(1)$,

and therefore X and Y are not statistically independent.

Example 7.7 The joint p.d. of two discrete r.v.'s X and Y is given by

$$f(x, y) = \frac{xy^2}{30} \qquad \text{for } x = 1, 2, 3 \text{ and } y = 1, 2.$$

Are X and Y independent?

The *marginal p.d.* for X is

$$g(x) = \sum_{y} f(x, y)$$

$$= \sum_{y=1}^{2} \frac{xy^2}{30} = \frac{x\,(1)^2}{30} + \frac{x\,(2)^2}{30} = \frac{x}{6}, \text{ for } x = 1, 2, 3;$$

and the *marginal p.d.* for Y is

$$h(y) = \sum_{x} f(x, y)$$

$$= \sum_{x=1}^{3} \frac{xy^2}{30} = \frac{1.y^2}{30} + \frac{2.y^2}{30} + \frac{3.y^2}{30} = \frac{y^2}{5}, \text{ for } y = 1, 2$$

Clearly, $\frac{xy^2}{30} = \frac{x}{6} \times \frac{y^2}{5}$, for $x = 1, 2, 3$ and $y = 1, 2$,

i.e. $f(x, y) = g(x) \cdot h(y)$

Hence X and Y are independent.

7.5.6 Continuous Bivariate Distributions. The *bivariate probability density function* of continuous r.v.'s X and Y is an integrable function $f(x, y)$ satisfying the following properties:

(i) $f(x, y) \geq 0$ for all (x, y).

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$, and

(iii) $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$.

The distribution function (d.f) of the bivariate r.v. (X, Y) is defined by

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du.$$

It should be noted that analogous to the relationship $\frac{d}{dx} F(x) = f(x)$,

we have $\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = f(x, y)$, wherever F is differentiable.

The *marginal* p.d.f. of the continuous r.v. X is

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

and that of the r.v. Y is

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

That is the marginal p.d.f. of any of the variables is obtained by integrating out the other variable from the joint p.d.f. between the limits $-\infty$ and $+\infty$.

The *conditional* p.d.f. of the continuous r.v. X given that Y takes the value y, is defined to be