Probability – Lecture 1

INTRODUCTION: BASIC IDEAS

Probability concerns unknown or uncertain outcomes. The concepts are widely applicable in many areas of physics as you will see as you go through the course. Here, we will deal with the basic ideas of probability.

A **probability** is the likelihood of a **random variable** taking on a specific value.

What is meant by random?

A random variable describes a value that corresponds to the outcome of a given experiment which is not known until *after* the experiment is conducted. It can be either **discrete** or **continuous**. An example of a discrete random variable would be number of people in Sheffield who wear size 9 shoes. An example of a continuous random variable would be the voltage produced by a failing 9 V battery which could be any value between 0 – 9 V.

Suppose that an event can occur in H ways from a total of N ways. Then the probability of occurrence of the event (its success) is denoted by:

Probability of success

$$p = \frac{H}{N}$$

and thus the probability that it will not occur (its failure) is given by q where

Probability of failure

$$q = \frac{N-H}{N} = 1 - \frac{H}{N} = 1 - p$$

Therefore, in situations where an event can either occur or not occur, the relation p + q = 1.

For example, we might want to know the probability of obtaining either a 3 or a 4 with a single toss of a die. There are 6 possible outcomes, i.e. the numbers 1 to 6. Therefore obtaining the numbers 3 or 4 represent 2/6 of the possible outcomes. The probability of getting a 3 or a 4 is therefore 1/3. And the probability of failing to get a 3 or a 4 is therefore 1 – 1/3 = 2/3.

CONDITIONAL PROBABILITY

If E_1 and E_2 are two events, the probability that E_2 occurs given that E_1 has already occurred is given by

$$P(E_2|E_1)$$

Is the outcome of an event determined in any way by the outcome of a previous event?

This is called conditional probability. If the occurrence of E_2 is not in any way determined by the previous occurrence of E_1, then we say that E_2 and E_1 are **independent**. Otherwise, these events are **dependent**.

We can also use the notation $P(E_1E_2)$ for the probability that both E_1 and E_2 will occur (compound event) for which

$$P(E_1E_2) = P(E_1)\cdot P(E_2).$$

An example would be the tossing of a coin several times. The probability that the 5th and 6th tosses of the coin resulted in "tails" would be:

$$P(E_1E_2) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

Independent

This is because the two events are *independent* – the probability of getting a "tails" on the 6th toss is in no way affected by obtaining "tails" on the 5th toss, and the probability of getting "tails" on the 5th toss is not in any way determined by the outcome of any of the previous tosses.

Another example in which the outcome of an event E_2 **is affected** by the outcome of a prior event E_1 is now given. We have a box containing 3 red balls and 2 black balls. E_1 represents "drawing the first ball as a black ball" and E_2 corresponds to "drawing the second ball as a black ball". The balls are not replaced after drawing. The probability of success in E_1 and E_2 (i.e. drawing the first two balls as black) is:

$$P(E_1E_2) = P(E_1)\cdot P(E_2|E_1)$$

The probability of the first ball drawn is black is clearly 2/5. However, if E_1 is a success, since this ball is not replaced the probability of drawing the second ball as black is now 1/4. Therefore the probability of achieving success for E_1 and E_2 is thus

$$P(E_1E_2) = \frac{2}{5} * \frac{1}{4} = \frac{1}{10}\ .$$

Dependent

We can see that if E_1 had failed (i.e. we had failed to draw a black ball initially), then the probability of drawing the second ball as black would be affected by this failure and would therefore be 2/4.

MUTUALLY EXCLUSIVE EVENTS

Two or more events are referred to as mutually exclusive if the occurrence of one event prevents the other event happening. If E_1

corresponds to "selecting a king from a deck of cards" and E_2 refers to "selecting a queen", then the probability of choosing a king **or** a queen in a single selection is given by $P(E_1+E_2)$ where

$$P(E_1+E_2)=P(E_1)+P(E_2)=\frac{4}{52}+\frac{4}{52}=\frac{8}{52}=\frac{2}{13}.$$

This is because E_1 and E_2 are **mutually exclusive** – in a single draw, selecting a king necessarily excludes the selection of a queen.

We can modify this example. Suppose now we say that E_1 is the event in which we "select a king" and E_2 is where we "select a spade". The probability of selecting a king or a spade is $P(E_1+E_2)$ where

$$P(E_1+E_2)=P(E_1)+P(E_2)-P(E_1E_2)$$
$$=\frac{4}{52}+\frac{13}{52}-\frac{1}{52}=\frac{16}{52}=\frac{4}{13}$$

This is because there are only 16 cards that fit the criteria (either a king or a spade) and not 17 as you may initially expect, since the King of Spades would represent a success in both E_1 and E_2. Thus these events are **not** mutually exclusive.

PROBABILITY DISTRIBUTIONS

A probability distribution gives the probability for each value of the random variable.

If we toss a pair of dice and let X symbolize the sum of the points obtained, then we can draw up a table (Table 5) showing the probability of obtaining each possible score. For instance, the probability of obtaining the score of 9 is 4/36 = 1/9 and so in 900 tosses of the dice, we would expect to obtain this score 100 times.

TABLE 5

X	2	3	4	5	6	7	8	9	10	11	12
p(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

The sum of all probabilities must equal 1.

We note that all the probabilities sum to 1, that is:

$$\Sigma\ p(X)=1$$

This means that the event of getting a score that is in the range $2 \le X \le 12$ is a **certainty** (there are no other possibilities). This type of probability distribution is known as a **discrete distribution** function, since only particular values of the random variable (the score) are possible. (In other words, a non-integer score e.g. 2.5 or a score lower than 2 or greater than 12 is impossible!)

Another type of probability distribution is a **continuous distribution** function in which the value of the random variable can take any value in a particular range. An example of this would be weight of each apple collected from one of the descendants of Newton's apple trees. Each apple would have a particular weight that would not be restricted in value like the score from the two dice above.

EXPECTED VALUE OR EXPECTATION

If p(S) is the probability that a person will receive a sum of money, the **expectation** is defined as p(S)S. Thus if the probability that a woman will receive £ 100,000 is 0.04 (i.e. if there is a probability of 4% that her wealthy partner may give her a gift of £ 100,000) then her expectation is £ 4,000. This is primarily a *mathematical concept* but has some use in physics.

BINOMIAL DISTRIBUTION

A special kind of probability distribution that is commonly found is the **binomial distribution**. This occurs in situations in which all outcomes of a trial must be classified into **two** categories. For instance, the trial of tossing a coin will result in either a head or a tail (two outcomes only) or a survey of the frequency of colour blindness will identify those who are colour blind and those who are not. Many statistical investigations have this "two-state" character. There are certain criteria that such a study must fit in order to be classified as a binomial experiment. These are:

1. **The investigation must have a fixed number of trials.**
2. **The trials must be independent (i.e. the outcome of each individual trial must not affect the outcome of any other trial).**
3. **Each trial must have outcomes that are classifiable into two categories only.**
4. **The probabilities for each outcome must remain constant throughout each trial.**

The formula for a binomial distribution.

$$p(X) = \frac{N!}{X!(N-X)!} p^X q^{N-X}$$

In this formula, p is the probability that an event will happen (a success) and q = 1-p is the probability that the event will fail to happen in any trial (a failure). The probability that the event will occur **exactly X times in N trials** (i.e. X successes and N-X failures) is given by p(X).

Key things to note.

The symbol "!" stands for "factorial"; thus N! means simply N(N-1)(N-2)…..1. So, for example:

6! = 6*5*4*3*2*1 = 720 and

9! = 9*8*7*6*5*4*3*2*1 = 362,880.

Also, by mathematical definition, 0! = 1 (this seems strange but is true by definition).

This expression seems difficult but it lets us calculate probabilities easily for investigations involving large numbers of trials – otherwise we would have to list all the possible combinations which would be time-consuming and tedious.

EXAMPLE

If we toss a coin 3 times, we might want to know the probability of obtaining exactly 2 tails. The tossing of a coin has two possible outcomes (heads or tails) and the outcome of any one toss plays no role in affecting the outcome of any other toss. Here, p = q = 0.5 (equal probability of getting heads or tails), the number of trials N = 3 and the number of successes (getting tails) = 2. Substituting these values into the formula above yields:

$$p(2) = \frac{3!}{2!(3-2)!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{3-2}$$
$$= \frac{6}{2} * \frac{1}{8} = \frac{6}{16} = \frac{3}{8}$$

As we would expect we get 3/8. We can check this by listing out the possible outcomes of the experiment:

HHH HHT HTH THH **TTH** **THT** **HTT** TTT

We can see that 3 of these combinations contain 2 tails exactly.

If we extended the number of trials to 6 tosses of the coin and asked how probable it would be to get exactly 3 tails, p(3) would be given by :

$$p(3) = \frac{6!}{3!(6-3)!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{6-3}$$
$$= \frac{720}{36}\left(\frac{1}{2}\right)^6 = \frac{720}{2304} = \frac{5}{16}$$

Listing out all the combinations of heads and tails for 6 tosses is a nightmare task, so this provides a much easier route to finding the required probability.

What does the distribution look like graphically?

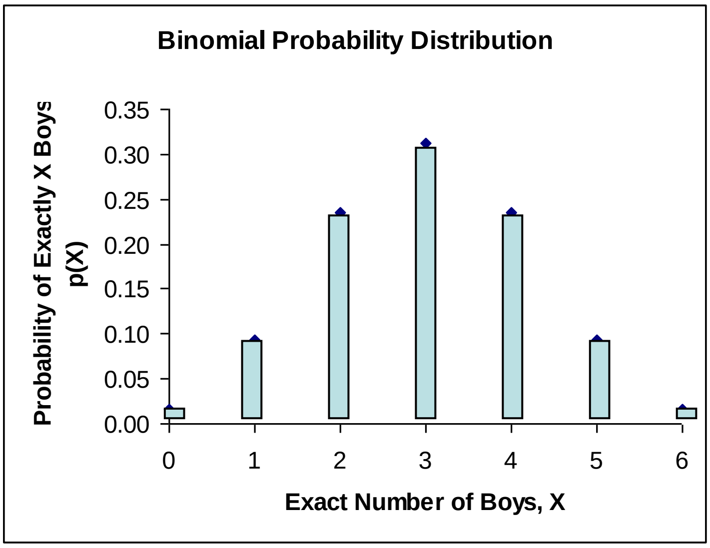
Let's consider another example and this time we will plot the data so that we can visualize the probability distribution. The Richardson family contains 6 children. We might want to know the probability that there are exactly 3 boys. We can use the formula above to show that p(3) = 5/16. We can find p(X) for all possible values of X, i.e. from 0 to 6 inclusive. These values are shown in Table 6 below.

TABLE 6

Exact No of Boys, X	0	1	2	3	4	5	6
p(X)	1/64	3/32	15/64	5/16	15/64	3/32	1/64

What is Σ p(X) ?

We notice that when we add all the probabilities together, we get 1. This makes sense since the event that the family contains either 0,1, 2, 3, 4, 5 or 6 boys is a certainty.



This graph highlights several characteristics of a binomial probability distribution. First we can see that in the case when p = q = 0.5 (i.e. the probability of the birth of a boy being equal to that of a girl) we obtain a symmetric distribution with the greatest probability occurring for X = 3. So it is most probable that the Richardson family will contain 3 boys (and 3 girls of course). You will notice that we haven't connected the data points with a continuous curved line. Instead we have used a series of bars as in a histogram. This is to emphasize that the binomial distribution is a *discrete* probability distribution. The exact number of boys can only take on discrete, specific values of 0, 1, 2, 3 etc.

Expected or mean value?

If we polled 1000 families each having 6 children, we would expect 1000 p(X) families to contain exactly X boys. So we would expect that 312.5 (in reality either 312 or 313) families to have 3 boys and 3 girls, 16 families to contain all boys or all girls and 234 families to contain either only 2 boys or only 2 girls etc.

We can change our language slightly to rephrase the above information. We can say that out of 1000 families, the mean number of families containing all boys or all girls is 16 or the mean number of families containing 3 boys and 3 girls is 312.5 etc.

Probability – Lecture 2

PROBABILITY DISTRIBUTION FUNCTIONS

A probability distribution function p(x) is a mathematical function which describes the probability of a random variable being equal to a particular value (of the random variable), x.

We can express the probability of the value of x lying in the range $x_1 \le x \le x_2$ as the integral:

$$\int_{x_1}^{x_2} p(x)dx$$

The probability of the random variable, x, being equal to a value of x somewhere within the full range of values covered by the probability function (x to x') is obviously 1 since we know that all the p(x) values should add to unity (see earlier definition for a discrete probability distribution function).

$$\int_{x}^{x'} p(x)dx = 1$$

EXAMPLE

Suppose we have the PDF (Probability Distribution Function)

$$f(x) = bx(5-x) \quad 0 \le x \le 5$$
$$f(x) = 0 \quad otherwise$$

then we can plot the PDF easily – see board.

Take notes on how to plot the function and insert plot of PDF here……..

We can calculate the value of the constant b using our integral definition above. Enter details below:

And then we can find any specific probability we like. For example, what is the probability of x being greater or equal to 2.3 and less than or equal to 4.5?

Finally, we can find the mean value using the following definition:

$$\bar{x} = \int_{x}^{x'} x f(x) dx$$

NORMAL DISTRIBUTION

The most common example of a continuous probability distribution is the **normal distribution**, sometimes referred to as the Gaussian distribution. Such a continuous distribution is required because in many statistical investigations the value of the random variable is not limited to discrete values as in the case of tossing a coin, throwing a die, giving birth etc. For example, we may be interested in investigating the birth weight of babies in a particular country. We may want to know the probability of a baby's weight, W, falling in the interval $6.25 \le W \le 6.5$ lb (pounds).

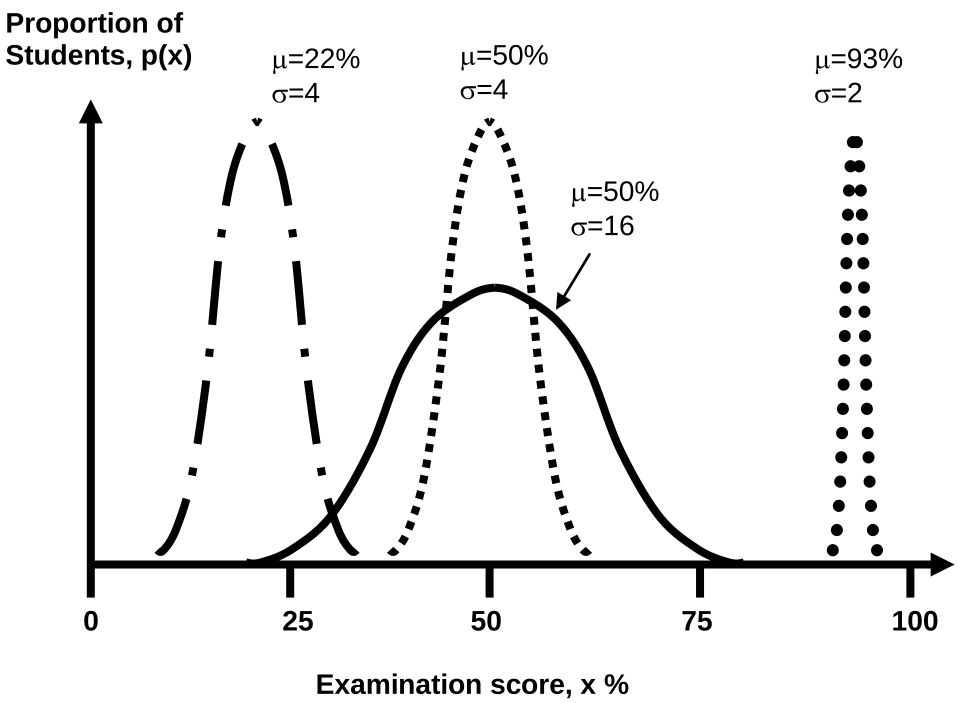
A continuous random variable has a normal distribution if the distribution is symmetric and bell-shaped and fits the equation given in the formula on the right of this box.

Again, mathematicians have helped us here and have developed a formula for such a distribution function:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

in which σ is the standard deviation, μ is the mean value, x is the value of the random variable, π is 3.14159… , e is 2.71828… and p(x) represents the probability of the value of the random variable being x.

The diagram below depicts several normal distributions representing mean scores in a fictitious physics examination. We can see the effect of changing the mean examination score and the standard deviation from the mean.

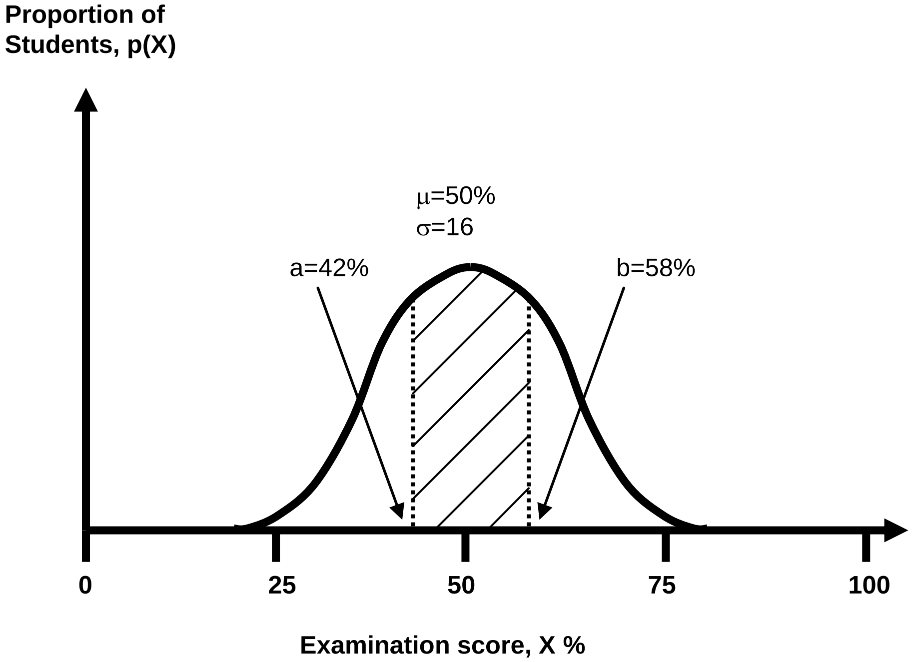


Interpretation of distributions

The breadth of each curve is a direct indication of the standard deviation – the broader the curve, the greater the standard deviation. If the mean average for the sample changes, then this corresponds to a shift in the bell-shaped curve – the mean corresponds to the x value at the centre of the distribution. In this example, the distribution with a mean of 93% represents a very easy examination in which a very large proportion of students scored highly and no students performed badly. The distributions with mean values of 50% represent two examinations. In one of these, the students' grades lay in a fairly tight range (approximately 37 – 62%) but in the other the spread of grades was greater (approximately 23 – 77%).

PROBABILITY AND AREA UNDER THE CURVE

Because the data plotted on the y-axis of a normal distribution graph always represents the probability of the random variable taking on a particular value x we can see that the area under the curve between two values of x (as indicated below) is a measure of the probability of x lying in the range $a \le X \le b$.



Taking this argument further reveals that the total area bounded by the curve and the x-axis must be a measure of the total probability of attaining a grade in the examination in the range $0 \le X \le 100$; in other words the total area under the curve represents **certainty**.

Probability – Lecture 3

MORE ON PROBABILITY DISTRIBUTION FUNCTIONS

In all cases, a particular function is a valid probability function if it integrates (over the entire range of values of random variable for which p(x) is non-zero) to give unity.

EXAMPLES

$$p(x) = ae^{-ax} \quad for\ x \geq 0$$
$$p(x) = 0 \quad for\ x < 0$$

See board for integral…….

Prove the function below is a valid PDF:

$$p(x) = \frac{3(a^2 - x^2)}{4a^3} \quad for - a \leq x \leq +a$$
$$p(x) = 0 \quad elsewhere$$

Sketch the PDF and by inspection, find the mean.

RETURN OF THE DISCRETE PDF

POISSON PDF

We now go back to the idea of discrete probability functions. Hopefully, you'll now understand the difference between continuous PDFs and discrete PDFs. The former described situations in which the random variable could take any value over a particular range (e.g. think back to our discussion of the heights of banana trees!!) whereas the latter described situations in which only certain possible values were available (e.g. the scores on dice).

Another useful discrete PDF can often be applied to situations in which we want to calculate the number of failures expected if we know the mean number of failures for a particular event.

The Poisson PDF is given by

$$p(x) = \frac{\lambda^n e^{-\lambda}}{n!}$$ for integer values of n = 0, 1, 2, 3……

The symbol λ above is defined as the mean. First of all, we should confirm that $\sum_0^\infty p(x) = 1$ (see board)

Now let's do an example…….

Certain microprocessors are known to have a failure rate of 1.8%. They are shipped in batches of 180. What is the probability that a batch has exactly two defective processors? (Assume Poisson PDF).

First of all we need to calculate the mean, λ.

This is simply the product of the number of number of processors and the expected failure rate = 3.24.

To find the probability of ***exactly*** 2 fails per batch, all we do is insert n=2 into the equation for p(x) (see board). You can calculate the probabilities of getting exactly a number of fails between 0 and 10 for practice. The table below shows the values:

Exact Number of Failures	**Value of n**	**p(x) – probability of obtaining exactly n failures per batch**
0	0	0.0392 = 3.92%
1	1	0.1269 = 12.69%
2	2	0.2056 = 20.56%
3	3	0.2220 = 22.20%
4	4	0.1798 = 17.98%
5	5	0.1165 = 11.65%
6	6	0.0629 = 6.29%
7	7	0.0291 = 2.91%
8	8	0.0118 = 1.18%
9	9	0.00042 = 0.42%
10	10	0.00014 = 0.14%

We may want to know the probability of obtaining less than 3 failures (i.e. 0, 1 or 2 failures). We find this probability simply by adding up the individual probabilities in the normal way:

$$p(<3) = p(0) + p(1) + p(2) = 0.372 = 37.2\%$$

You should notice that if you add up all the percentage values, you obtain 99.94%. This means that nearly all the probability of achieving n failures lies in the region 0 – 10. This sort of information is very useful to manufacturers since they can bargain on achieving an absolute minimum of 170 working processors per batch of 180 and so can plan their sales budgets etc. on these figures – this would of course be a very conservative estimate, since you can see that the total probability of obtaining 4 or less failures is >75% for example.

That's about it for probability this year.