

High School Mathematics Extensions/Discrete Probability

Introduction

Probability theory is one of the most widely applicable mathematical theories. It deals with uncertainty and teaches you how to manage it. It is simply one of the most useful theories you will ever learn.

Please do not misunderstand: We are not learning to predict things; rather, we learn to utilise predicted *chances* and make them useful. Therefore, we don't care about questions like *what is the probability it will rain tomorrow?*, but given that the probability is 60% we can make deductions, the easiest of which is *the probability it will not rain tomorrow is 40%*.

As suggested above, a *probability* is a percentage, and it's between 0% and 100% (inclusive). Mathematicians like to express a *probability* as a *proportion*, i.e. as a number between 0 and 1. So the probability that it will rain tomorrow is 0.6.

Application

You might ask why we are even studying probability. Let's see a very quick example of probability in action.

Consider the following gambling game: Toss a coin; if it's heads, I give you \$1; if it's tails, you give me \$2. You will easily notice that it is not a fair game - the chances are the same (50%-50%) but the rewards are different. Even though we are playing with probability, there are useful, and sometimes not so obvious, conclusions we can make: one of them is that *in the long run* I will become richer and you will become poorer.

One More Real Life Example: I observed one day that there are dark clouds outside. So I asked myself, should I bring an umbrella? I use my observation of dark clouds(probability) as per my usual daily deciding routine. Since in past experiences, dark clouds are early warning signs of rain, I am more likely to bring an umbrella.

In real life, probability theory is heavily used in risk analysis by economists, businesses, insurance companies, governments, etc. An even wider usage is its application as the basis of statistics, which is the main basis of all scientific research. Two branches of physics have their bases tied in probability. One is clearly shown from its name: statistical mechanics. Another branch is quantum physics.

Why discrete probability?

Probability comes in two flavours: discrete and continuous. The continuous case is considered to be far more difficult to understand, and much less intuitive, than discrete probability, and it requires knowledge of calculus. But we will touch on a little bit of the continuous case later on in the chapter.

Event and Probability

Roughly, an *event* is something we can assign a *probability* to. For example *the probability it will rain tomorrow is 0.6*; here, the event is *it will rain tomorrow*, and the assigned probability is 0.6. We can write

$$P(\text{it will rain tomorrow}) = 0.6$$

Mathematicians typically use abstract letters to represent events. In this case we choose A to represent the event *it will rain tomorrow*, so the above expression can be written as

$$P(A) = 0.6$$

Another example is *a fair die will turn up 1, 2, 3, 4, 5 or 6 with equal probability each time it is tossed*. Let B be the event that it turns up 1 in the next toss. We write:

$$P(B) = 1/6$$

Misconception

Please note that the probability $1/6$ does **not** mean that it will turn up 1 in at most six tries. Its precise meaning will be discussed later on in the chapter. Roughly, it just means that on the long run (i.e. the die being tossed a large number of times), the proportion of 1s will be very close to $1/6$.

Impossible and certain events

Two types of events are special. One type are the impossible events (e.g., a roll of a die will turn up 7); the other type are certain to happen (e.g., a roll of a die will turn up as one of 1, 2, 3, 4, 5 or 6). The probability of an impossible event is 0, while that of a certain event is 1. We write

$$P(\text{Impossible event}) = 0$$

$$P(\text{Certain event}) = 1$$

The above reinforces a very important principle concerning probability. Namely, the range of probability is between 0 and 1. You can **never** have a probability of 2.5! So remember the following

$$0 \leq P(E) \leq 1$$

for all events E.

Complement of an event

A most useful concept is the **complement** of an event. Here we use $\overline{E}$ to represent the *event that the die will NOT turn up 1 in the next toss*. Generally, putting a bar over a variable (that represents an event) means the opposite of that event. In the above case of a die:

$$P(\overline{E}) = 5/6$$

it means *The probability that the die will turn up 2, 3, 4, 5 or 6 in the next toss is 5/6*. Please note that

$$P(\overline{E}) = 1 - P(E)$$

for any event E.

There are some other notations for (ways to write) **complement** rather than putting a bar (line) on top: prime (A') and star (A^*). Both A' and A^* mean: $\overline{A}$

Combining independent probabilities

Independent probabilities can be combined to yield probabilities for more complex events. I stress the word *independent* here, because the following demonstrations will not work without that requirement. The exact meaning of the word will be discussed a little later on in the chapter, and we will show why *independence* is important in Exercise 10 of this section.

Adding probabilities

Probabilities are added together whenever a single event can occur in multiple "ways". As this is a rather loose concept, the following example may be helpful. Consider rolling a single die; if we want to calculate the probability for, say, rolling an odd number, we must add up the probabilities for all the "ways" in which this can happen -- rolling a 1, 3, or 5. Consequently, we come to the following calculation:

$$P(\text{rolling an odd number}) = P(\text{rolling a 1}) + P(\text{rolling a 3}) + P(\text{rolling a 5}) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2 = 0.5$$

Note that the addition of probabilities is often associated with the use of the word "or" -- whenever we say that some event E consist of the events X, Y, **or** Z (being satisfied if **any** of the events occur) we use addition to combine their probabilities (if they are disjoint, see below).

A general rule of thumb is that the probability of an event and the probability of its complement must add up to 1. This makes sense, since we intuitively believe that events, when well-defined, must either happen or not happen.

Multiplying probabilities

Probabilities are multiplied together whenever an event occurs in multiple "stages" or "steps." For example, consider rolling a single die twice; the probability of rolling a 6 in two consecutive rolls (two times back to back) is calculated by multiplying the probabilities for the individual steps involved since the two events are independent. Intuitively, the first step is simply the first roll, and the second step is the second roll. Therefore, the final probability for rolling a 6 twice is as follows:

$$P(\text{rolling a 6 twice}) = P(\text{rolling a 6 the first time}) \times P(\text{rolling a 6 the second time}) = \frac{1}{6} \times \frac{1}{6} = 1/36 \approx 0.028$$

(or 2.8%)

Similarly, note that the multiplication of probabilities is often associated with the use of the word "and" -- whenever we say that some event E is equivalent to **all** of the events X, Y, **and** Z occurring, we use multiplication to combine their probabilities (if they are independent).

Also, it is important to recognize that the product of multiple probabilities must be less than or equal to each of the individual probabilities, since probabilities are restricted to the range 0 through 1. This agrees with our intuitive notion that relatively complex events are usually less likely to occur.

Combining addition and multiplication

It is often necessary to use both of these operations simultaneously. Once again, consider a die being rolled twice in succession. In contrast with the previous case, we will now consider the event of rolling two numbers that add up to 3. In this case, there are clearly two steps involved, and therefore multiplication will be used, but there are also multiple ways in which the event under consideration can occur, meaning addition must be involved as well. The die could turn up 1 on the first roll and 2 on the second roll, or 2 on the first and 1 on the second. This leads to the following calculation:

$$P(\text{rolling a sum of 3}) = P(1 \text{ on 1st roll}) \times P(2 \text{ on 2nd roll}) + P(2 \text{ on 1st roll}) \times P(1 \text{ on 2nd roll}) = \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} = 1/18 \approx 0.056 \text{ (or 5.6\%)}$$

This is only a simple example, and the addition and multiplication of probabilities can be used to calculate much more complex probabilities.

Exercises

Let A represent the number that turns up in a (fair) die roll, let C represent the number that turns up in a separate (fair) die roll, and let B represent a card randomly picked out of a deck:

1. A die is rolled. What is the probability of rolling a 3 i.e. calculate $P(A = 3)$?
2. A die is rolled. What is the probability of rolling a 2, 3, **or** 5, i.e. calculate $P(A = 2, 3 \text{ or } 5)$?
3. What is the probability of choosing a card of the suit Diamonds (in a 52-card deck)?
4. A die is rolled and a card is randomly picked from a deck of cards. What is the probability of rolling a 4 **and** picking the Ace of Spades, i.e. calculate $P(A = 4) \times P(B = \text{Ace of spades})$.
5. Two dice are rolled together. What is the probability of getting a 1 and a 3?
6. Two dice are rolled separately. What is the probability of getting a 1 and a 3, regardless of order?
7. Calculate the probability of rolling two numbers that add up to 7.

8. (Optional) Let C be the number rolled on the first die and A be the number rolled on the second die. Show that the probability of C is equal to A is $1/6$.

9. Let C and A be as in exercise 8. What is the probability that C is greater than A ?

10. Gareth was told that in his class 50% of the pupils play football, 30% play video games and 30% study mathematics. So if he was to choose a student from the class randomly, he calculated the probability that the student plays football, plays video games, and studies mathematics is $50\% + 30\% + 30\% = 1/2 + 3/10 + 3/10 = 11/10$. But all probabilities should be between 0 and 1. What mistake did Gareth make?

Solutions

1. $P(A = 3) = 1/6$

2. $P(A = 2) + P(A = 3) + P(A = 5) = 1/6 + 1/6 + 1/6 = 1/2$

3. $P(B = \text{Ace of Diamonds}) + \dots + P(B = \text{King of Diamonds}) = 13 \times 1/52 = 1/4$

4. $P(A = 4) \times P(B = \text{Ace of Spades}) = 1/6 \times 1/52 = 1/312$

5. $P(A = 1) \times P(A = 3) + P(A = 3) \times P(A = 1) = 1/36 + 1/36 = 1/18$

6. $P(A = 1) \times P(A = 3) + P(A = 3) \times P(A = 1) = 1/36 + 1/36 = 1/18$ This is the same answer as the problem above because in both cases the outcome for each individual die remains independent of the other regardless of whether or not they are thrown simultaneously. Another way of calculating the same answer is to consider that the first dice can be a one or a three but the second can only be one number - the opposite of the first dice, i.e. a 3 if the first dice was 1, or a 1 if the first dice was 3. That gives: $P(A=1 \text{ or } A=3) \times P(\text{opposite}) = 2/6 \times 1/6 = 2/36 = 1/18$.

7. Here are the possible combinations: $1 + 6 = 2 + 5 = 3 + 4 = 7$. Probability of getting each of the combinations are $1/18$ as in exercise 6. There are 3 such combinations, so the probability is $3 \times 1/18 = 1/6$.

8. $P(A=C) = P(1,1) + P(2,2) + \dots + P(6,6) = 6 \times 1/36 = 1/6$

9. Since both dice are fair, $C > A$ is just as likely as $C < A$. So

$$P(C > A) = P(C < A) = X$$

and

$$P(C > A) + P(C < A) + P(A = C) = 1$$

But

$$P(A = C) = 1/6$$

$$\text{so } P(C > A) + P(C < A) = 5/6$$

$$2X = 5/6$$

$$P(C > A) = X = 5/12.$$

OR

$$P(A=1) \times P(C=2,3,4,5,6) + P(A=2) \times P(C=3,4,5,6) + P(A=3) \times P(C=4,5,6) + P(A=4) \times P(C=5,6) + P(A=5) \times P(C=6)$$

$$1/6 \times 5/6 + 1/6 \times 4/6 + 1/6 \times 3/6 + 1/6 \times 2/6 + 1/6 \times 1/6$$

$$1/6 \times (5+4+3+2+1)/6$$

$$5/12$$

10. These three sets overlap so, for example, to get the probability of someone belonging to all three sets, you need to multiply (assuming they are independent), not add. $P(F \text{ and } V \text{ and } M) = .5 \times .3 \times .3 = 0.045$. It is necessary to remember that the events of playing football, playing video game, studying mathematics, or being human, a male, living in the USA, etc are all possible. Although the likelihood and independence of these events/states may be debatable, the fact that the probability of any strange combination must be less than one must hold.

Random Variables

A *random experiment*, such as *throwing a die* or *tossing a coin*, is a process that produces some uncertain outcome. We also require that a random experiment can be repeated easily. In this section we shall start using a capital letter to represent the outcome of a random experiment. For example, let D be the outcome of a die roll. D could take the value 1, 2, 3, 4, 5 or 6, but it is uncertain. We say D is a *discrete random variable*. Suppose now that I throw a die and it turns up 5. We say the *observed value* of D is 5.

A random variable is simply the outcome of a certain random experiment. It is usually denoted by a CAPITAL letter, but its observed value is not. For example let

$$D_1, D_2, \dots, D_n$$

denote the outcome of n die throws, then we usually use

$$d_1, d_2, \dots, d_n$$

to denote the observed values of each D_i .

From here on, random variable may be abbreviated as "rv" (a common abbreviation in other probability texts).

The Bernoulli experiment

(This section is optional and it assumes knowledge of binomial expansion.)

A coin-toss is a simpler, specific form of the Bernoulli experiment. If we toss a coin, we will expect to get a head or a tail equally probably. A Bernoulli experiment is slightly more versatile than that, in that the two possible outcomes need not have the same probability.

In a Bernoulli experiment you will either get a

success, denoted by 1, with probability p (where p is a number between 0 and 1)

or a

failure, denoted by 0, with probability $1 - p$.

If the random variable B is the outcome of a Bernoulli experiment, and the probability of a successful outcome of B is p , we say B comes from a *Bernoulli distribution* with success probability p (where $X \sim D$ means that the random variable X has the probability distribution D):

$$B \sim \text{Ber}(p)$$

For example, if

$$C \sim \text{Ber}(0.65)$$

then

$$P(C = 1) = 0.65$$

and

$$P(C = 0) = 1 - 0.65 = 0.35$$

Binomial Distribution

If we repeat a Bernoulli experiment n times and count the number of successes, we get a binomial distribution. For example:

$$C_i \sim \text{Ber}(p)$$

for $i = 1, 2, \dots, n$. That is, there are n variables $C_1, C_2, \dots, C_n$ and they all come from the same Bernoulli distribution.

We consider:

$$B = C_1 + C_2 + \dots + C_n$$

, then B is simply the random variable that counts the number of successes in n trials (experiments). Such a variable is called a binomial variable, and we write

$$B \sim \text{Bin}(n, p)$$

Example 1

Aditya, Gareth, and John are equally able. Their probability of scoring 100 in an exam follows a Bernoulli distribution with success probability 0.9. What is the probability of

- i) Only one of them getting 100?
- ii) Two of them getting 100?
- iii) All 3 getting 100?
- iv) None getting 100?

Solution

We are dealing with a binomial variable, which we will call B . And

$$B \sim \text{Bin}(3, 0.9)$$

i) Aditya's (as well as John and Gareth's) probability of scoring 100 is 0.9 or 90%. We can write this as

$$P(S = 100) = 0.9$$

... where S represent the score of any of them. The probability of any of them getting 100 (success) and the other two getting below 100 (failure) is

$$0.9 \times 0.1 \times 0.1 = 0.009$$

but there are 3 possible candidates for getting 100, so

$$P(B = 1) = 3 \times 0.009 = 0.027$$

ii) We want to calculate

$$P(B = 2)$$

The probability is

$$0.9 \times 0.9 \times 0.1 = 0.081$$

but there are $\binom{3}{2}$ combinations of candidates for getting 100, so

$$P(B = 2) = \binom{3}{2} \times 0.081 = 0.243$$

iii) To calculate

$$P(B = 3) = 0.9 \times 0.9 \times 0.9 = 0.729$$

iv) The probability of "None getting 100" is getting 0 success, so

$$P(B = 0) = 0.1 \times 0.1 \times 0.1 = 0.001$$

The above example strongly hints at the fact the binomial distribution is connected with the binomial expansion. The following result regarding the binomial distribution is provided without proof; the reader is encouraged to check its correctness.

If

$$B \sim \text{Bin}(n, p)$$

then

$$P(B = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

This is the k th term of the binomial expansion of $(p + q)^n$, where $q = 1 - p$.

Events

In the previous sections, we have slightly abused the word "event". An event should be thought of as a collection of random outcomes of a certain random variable.

Let us introduce some notation first. Let A and B be two events, we define

$$A \cap B$$

to be the event of A and B . We also define

$$A \cup B$$

to be the event of A or B . As demonstrated in exercise 10 above,

$$P(A \cup B) \neq P(A) + P(B)$$

in general.

Let's see some examples. Let A be the event of getting a number less than or equal to 4 when tossing a die, and let B be the event of getting an odd number. Now

$$P(A) = 2/3$$

and

$$P(B) = 1/2$$

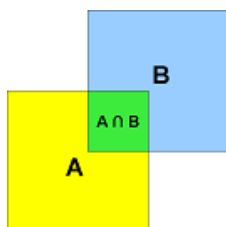
but the probability of A or B does not equal to the sum of the probabilities:

$$P(A \cup B) \neq P(A) + P(B) = \frac{1}{2} + \frac{2}{3} = \frac{7}{6}$$

as $7/6$ is greater than 1.

It is not difficult to see that the event of throwing a 1 or 3 is included in both A and B . So if we simply add $P(A)$ and $P(B)$, some events' probabilities are being added twice.

The Venn diagram below should clarify the situation a little more,



Think of the blue square as the probability of B and the yellow square as the probability of A . These two probabilities overlap, and the space where they overlap is the probability of A and B . So the probability of A or B should be:

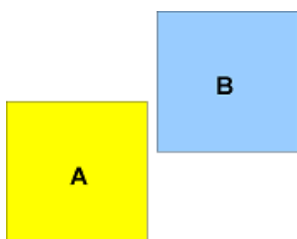
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The above formula is called the *Simple Inclusion-Exclusion Principle*.

If for events A and B , we have

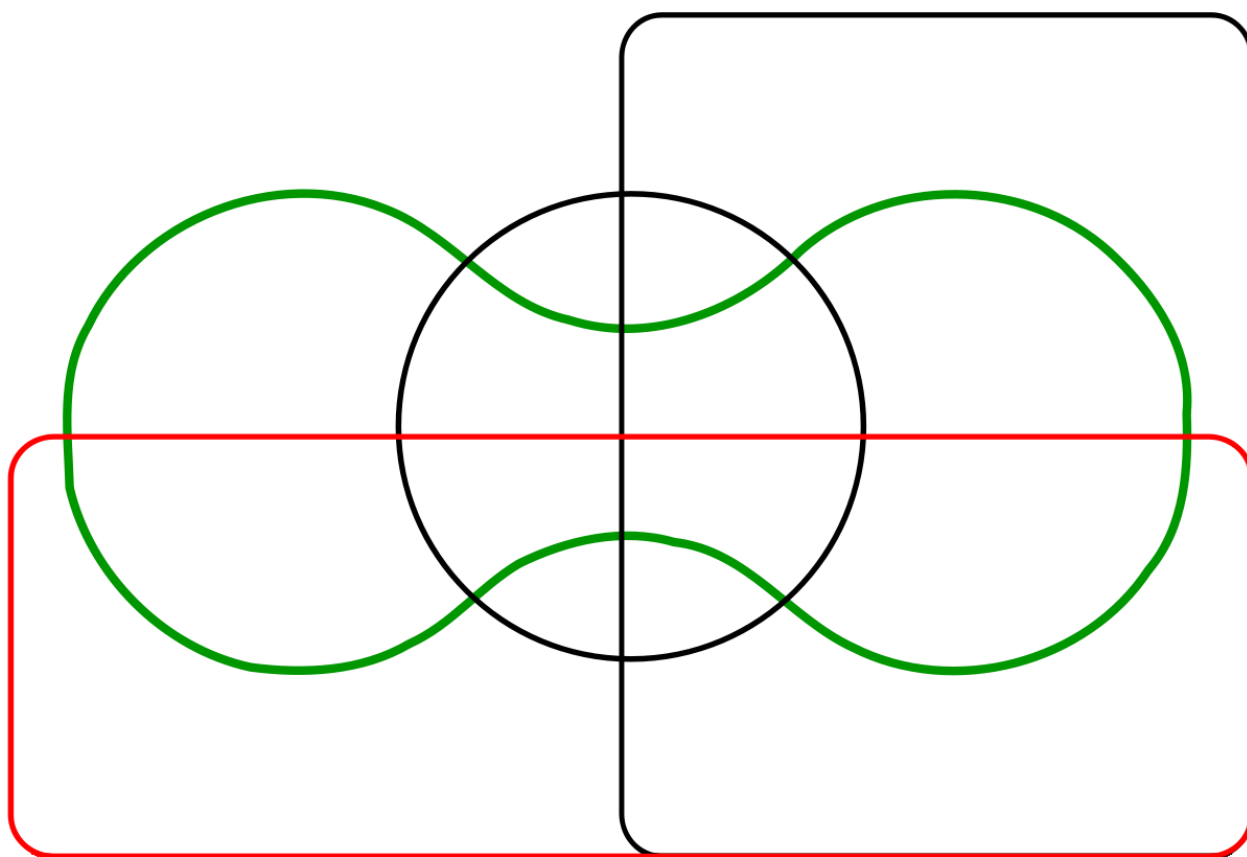
$$P(A \cap B) = 0$$

we say A and B are **disjoint**. The word means *to separate*. If two events are disjoint, the following Venn diagram represents them:



Info—Venn diagram

Traditionally, Venn diagrams are used to illustrate sets graphically. A set is simply a collection of things -- for instance, $\{1, 2, 3\}$ is a set consisting of 1, 2 and 3. Venn diagrams are usually drawn round. It is generally very difficult to draw Venn diagrams for more than 3 intersecting sets. To demonstrate why, here is a Venn diagram showing four intersecting sets:



Expectation

The expectation of a random variable can be roughly thought of as the long-term average of the outcome of a certain repeatable random experiment, where by *long-term average* we mean that we perform the underlying experiment many times and average the outcomes. For example, let D be as above; the observed values of D (1, 2 ... or 6) are equally likely to occur. So if you were to toss the die a large number of times, you would expect each of the numbers to turn up roughly an equal number of times. So the expectation is

$$\frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

We denote the expectation of D by $E(D)$, so

$$E(D) = 3.5$$

We should now properly define the expectation.

Consider a random variable R , and suppose the possible values it can take are $r_1, r_2, r_3, \dots, r_n$. We define the expectation to be

$$E(R) = r_1 P(R = r_1) + r_2 P(R = r_2) + \dots + r_n P(R = r_n)$$

Think about it: Taking into account the expectation is the long term average of the outcomes. Can you explain why $E(R)$ is defined the way it is?

Example 1 In a fair coin toss, let 1 represent tossing a head and 0 a tail. The same coin is tossed 8 times. Let C be a random variable representing the number of heads in 8 tosses. What is the expectation of C , i.e. calculate $E(C)$?

Ans. $E(C) = \sum [r \times P(C=r)]$ where $0 \leq r \leq 8$

$$\begin{aligned} P(r) &= \binom{8}{r} (1/2)^r \times (1-1/2)^{8-r} = \binom{8}{r} (1/2)^8 \\ E(C) &= 0 \times \binom{8}{0} (1/2)^8 + 1 \times \binom{8}{1} (1/2)^8 + \dots + 8 \times \binom{8}{8} (1/2)^8 \\ &= (0 + 8 + 56 + 168 + 280 + 280 + 168 + 56 + 8) \times (1/2)^8 \\ &= 1024 \times (1/256) \\ &= 4 \end{aligned}$$

So the expectation value is 4

Areas as probability

The uniform distributions...

Order Statistics

Estimate the x in $U[0, x]$

Addition of the Uniform distribution

Adding $U[0,1]$'s and introduce the CLT.

...CLT - **Central Limit Theorem:** In any set of sample distributions, as the number of samples taken increases, the overall mean distribution of the sample distributions will approach a Normal distribution.

The CLT is important in Statistical inference where small samples are taken of entire populations to draw conclusions on the entire population.

Feedback

What do you think? Too easy or too hard? Too much information or not enough? How can we improve? Please let us know by leaving a comment in the discussion section. Better still, edit it yourself and make it better.

Article Sources and Contributors

High School Mathematics Extensions/Discrete Probability *Source:* <http://en.wikibooks.org/w/index.php?oldid=2543721> *Contributors:* 2602:306:CC7F:1590:DD2:150A:C9B9:D9B2, Abramsky, Adrignola, AllenZh, ArielGenesis, Ars Tapatun, Ashokkrjha, Bobbitjl, Buggi22, Dielson Sales, Guanaco, Inkling, Jafeluv, Jfmantis, Jithrae, Jomegat, Jonathanjaffe, Jwwalker, Khola07, Lemontea, LivingFont, Mark Simonyi, Masinghal, Mickoz, Mike.lifeguard, Nijdam, Nothing, Pozyton, QuiteUnusual, Recent Runes, Spongebob88, Xania, Xiaodai, Ysaimanoj, 107 anonymous edits

Image Sources, Licenses and Contributors

image:HSE_Venn1.png *Source:* http://en.wikibooks.org/w/index.php?title=File:HSE_Venn1.png *License:* GNU Free Documentation License *Contributors:* Az1568, Webaware, Xiaodai
image:HSE_Venn_2.png *Source:* http://en.wikibooks.org/w/index.php?title=File:HSE_Venn_2.png *License:* GNU Free Documentation License *Contributors:* Az1568, Webaware, Xiaodai
Image:Edwards-Venn-four.svg *Source:* <http://en.wikibooks.org/w/index.php?title=File:Edwards-Venn-four.svg> *License:* GNU Free Documentation License *Contributors:* User:HB
File:Narender.jpg *Source:* <http://en.wikibooks.org/w/index.php?title=File:Narender.jpg> *License:* Creative Commons Attribution-Sharealike 3.0 *Contributors:* User:Khola07

License

Creative Commons Attribution-Share Alike 3.0 Unported
[//creativecommons.org/licenses/by-sa/3.0/](http://creativecommons.org/licenses/by-sa/3.0/)