Probability

Why Study Probability?

Probability theory is an important branch of mathematics useful in daily life decision making as well as all scientific and engineering endeavors. It is a fundamental way of thinking and problem solving skill.

“We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it. … It is remarkable that this science, which originated in the consideration of games of chance, should have become the most important object of human knowledge. … The most important questions of life are, for the most part, really only problems of probability.” Simon de Laplace (1749-1827)

Definition of Probability

"To study a phenomenon, one must reduce all events of the same type to a certain number of equally possible cases, and then the probability is a fraction whose numerator represents the number of cases favorable to the event and whose denominator represents the number of possible cases." Simon de Laplace (1749-1827), *Théorie Analytique des Probabilités*

$$P(event) = \frac{number\ of\ favorable\ cases}{number\ of\ possible\ cases}$$

Interpretation of Probability

- Frequency interpretation: through repeated experiments, the probability of the outcome is the proportion of the experiments that result in the outcome. Usually used by scientists

  e.g. P(a fair die will roll a 6) = $\frac{1}{6}$

  You can repeat experiment a lot of times to get the number 1/6

- Subjective interpretation: a belief of the chance that an outcome will happen. Usually used by philosophers or decision maker

  e.g. P(person x will become the chief executive) = $\frac{1}{3}$

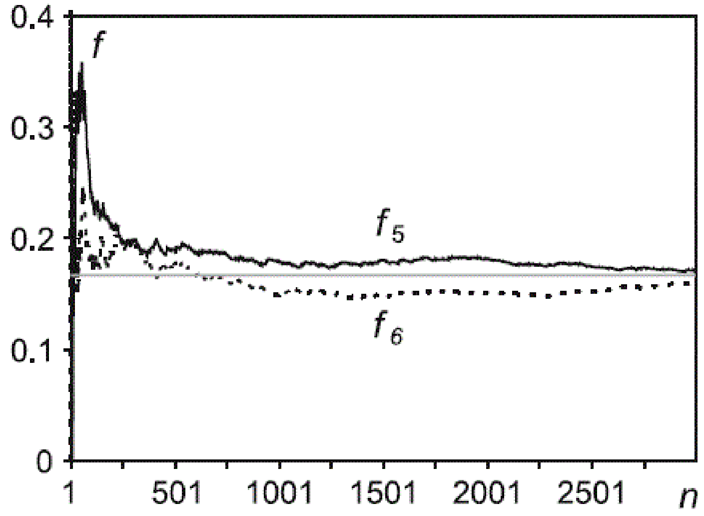
  The event either occurs or not occur. You cannot repeat the experiment

Example

How would you interpret the following using the frequency interpretation and the subjective interpretation?

$$P(tomorrow\ will\ be\ a\ rainy\ day) > \frac{1}{5}$$

$$P(a\ 5\ and\ then\ a\ 6\ is\ obtained\ in\ two\ dice\ rolls) = \frac{1}{36}$$



Frequency curves of throwing '5' and '6' in a real dice

Example: (It is important to divide into equally possible cases)

Consider a family with two children. Assume

$$P(boy) = P(girl) = 0.5$$

What is the probability of the family with one boy and one girl?

One may intuitively divide into three cases:
Case A: the family has two boys
Case B: the family has one boy and one girl
Case C: the family has two girls

Then $P(one\ boy\ and\ one\ girl) = \frac{1}{3}$ which is wrong! Why?

One should divide into equally possible cases, or elementary events. These are events with equal probability to occur.

{MM, MF, FM, FF}

M: Male F: Female

Hence $P(one\ boy\ and\ one\ girl) = \frac{1}{2}$

(Name a hidden assumption behind this model)

In general, we need to have a language to reason about events and their interaction with each other. Below is such a language

Sample Space and Events

- Sample Space S : The set of all possible outcomes of an experiment

  E.g. The sample space of a roll of fair dice is

  $S = \{1, 2, 3, 4, 5, 6\}$

- Event E : Any subset of the sample space. An event is a set consisting of possible outcomes of an experiment.

  E is the event that a dice roll gives an even number

  $E = \{2, 4, 6\}$

- Union of two events $E \cup F$ or

  F is the event of all dice roll gives a number larger than 3

  $F = \{4, 5, 6\}$

  $E \cup F$ is the event that "the dice roll gives an even number or a number larger than 3"

  $E \cup F = \{2,4,5,6\}$

  $P(E \cup F) = P(E) + P(F) - P(EF)$

- Intersection of two events $E \cap F$ and

  $E \cap F$ is the event that "the dice roll gives an even number larger than 3"

  $E \cap F = \{4,6\}$

  Sometimes, we write $E \cap F$ as EF

- Complement of an event : The complement of an event E, denoted E^c, consists of all outcomes in the sample space S that are not in E

  $E^c = \{1,3,5\}$

  As another example, if $E = \{1\}$ (the event that the dice rolls '1'), then $E^c = \{2,3,4,5,6\}$ (the event that the dice does not roll '1')

- E is contained in F if any outcome in E is also an outcome in F. We denote this as $E \subseteq F$ (In the book, $\subset$ is used)

  If $E = \{2,4,6\}$ and $F = \{2,3,4,5,6\}$, $E \subseteq F$

- E and F are equal if $E \subseteq F$ and F $\supseteq E$

- Mutually exclusive events : Two events E and F are mutually exclusive if $E \cap F = \emptyset$ ($\emptyset$ denotes empty set)

  E : the dice roll gives an even number $E = \{2,4,6\}$

  F : the dice roll gives an odd number $F = \{1,3,5\}$

  $E \cap F = \phi$

- We may denote more than one events using similar notations, e.g.

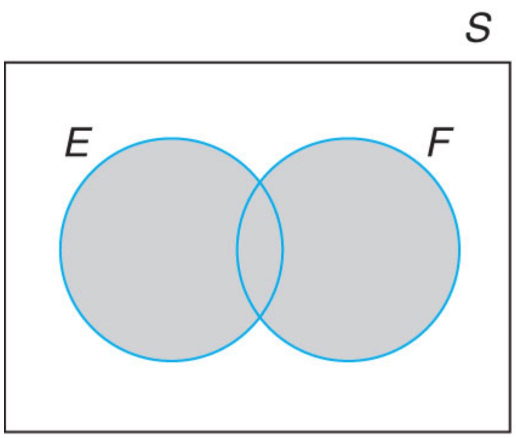
  $\bigcup_1^n E_i = E_1 \cup E_2 \cup \cdots \cup E_n$

  $\bigcap_1^n E_i = E_1 \cap E_2 \cap \cdots \cap E_n$

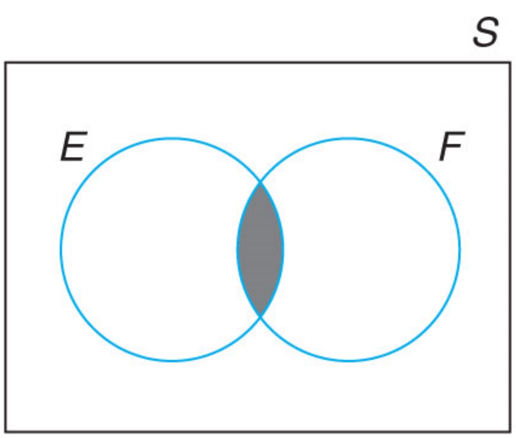
Venn Diagram

Invented by the 19th century English mathematician John Venn

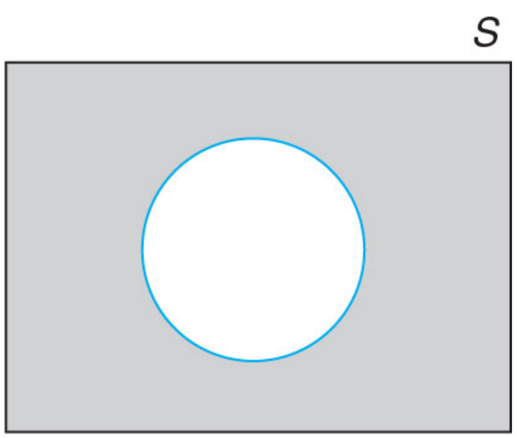
Represent events and their interactions by circles. We can use it to prove various theorems about events.



(a) Shaded region: $E \cup F$



(b) Shaded region: EF



(c) Shaded region: E^c

Theorems about Events

These theorems are common to both probability theory and logic theory

Commutative Law $E \cup F = F \cup E \quad EF = FE$

Associative Law $(E \cup F) \cup G = E \cup (F \cup G)$

$(EF)G = E(FG)$

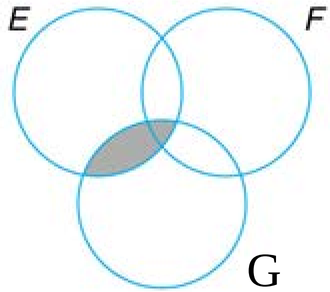
Distributive Law $(E \cup F)G = \mathrm{EG} \cup FG$

$EF \cup G = (E \cup G)(F \cup G)$

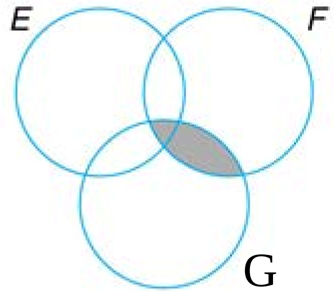
DeMorgan's Laws $(E \cup F)^c = E^c F^c$

$(EF)^c = E^c \cup F^c$

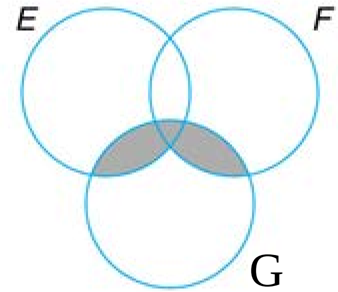
Example: Proving distributive law $(E \cup F)G = \mathrm{EG} \cup FG$



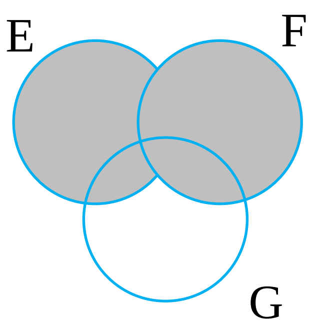
EG



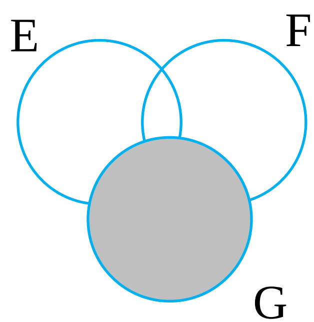
FG



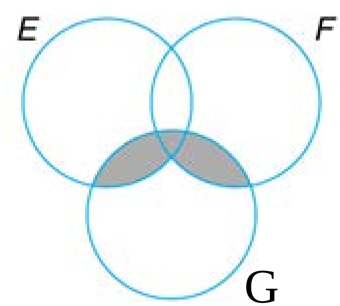
$\mathrm{EG} \cup FG$



$E \cup F$



G



$(E \cup F)G$

Another way to prove the law is by enumerating all cases using truth tables

E	F	G	EG	FG	$E \cup F$	EG $\cup FG$	$(E \cup F)G$
T	T	T	T	T	T	T	T
T	T	F	F	F	T	F	F
T	F	T	T	F	T	T	T
T	F	F	F	F	T	F	F
F	T	T	F	T	T	T	T
F	T	F	F	F	T	F	F
F	F	T	F	F	F	F	F
F	F	F	F	F	F	F	F

same truth values

Axioms of Probability

- One can derive the probability theory from the axioms below:

- AXIOM 1 $0 \leq P(E) \leq 1$
- AXIOM 2 $P(S) = 1$
- AXIOM 3

  For any sequence of mutually exclusive events $E_1, E_2, \ldots$

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Using Axioms to Prove Propositions

Proposition: $P(E^c) = 1 - P(E)$

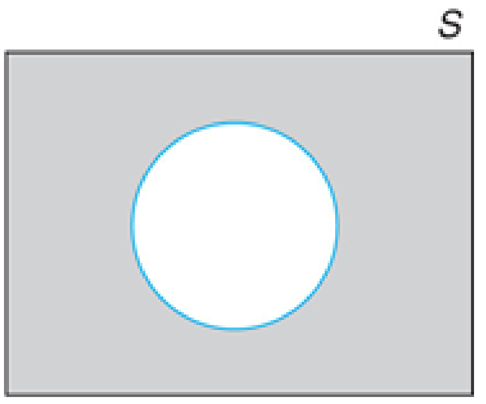
Proof:

By Axiom 2, $1 = P(S)$

Since $S = \mathrm{E} \cup E^c$, $P(S) = P(E \cup E^c)$

By Axiom 3, $P(E \cup E^c) = P(E) + P(E^c)$

Hence $P(E^c) = 1 - P(E)$



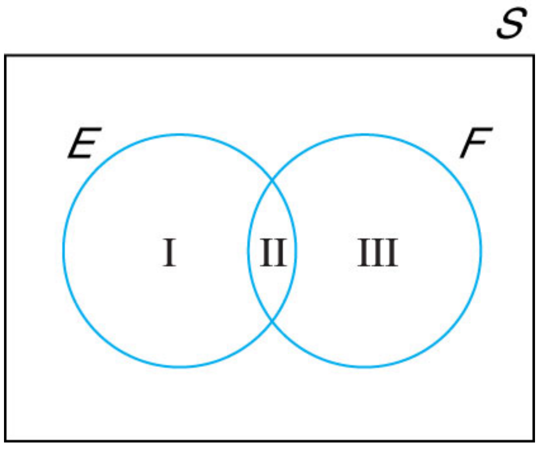
shaded region is E^c

Proposition: $P(E \cup F) = P(E) + P(F) - P(EF)$

(inclusion-exclusion principle)

Intuition: because region II is counted twice

Proof:



As regions I, II, III are mutually exclusive. By Axiom 3

$P(E \cup F) = P(I) + P(II) + P(III)$

$P(E) = P(I) + P(II)$

$P(F) = P(II) + P(III)$

Thus

$P(E \cup F) = P(E) + P(F) - P(II) = P(E) + P(F) - P(EF)$

Example

In a roll of fair dice, let E be the event the number is even. Let F be the event that the number is less than or equal to 4. What is the probability that the number is even or less than or equal to 4?

$$E = \{2, 4, 6\}$$
$$F = \{1,2,3,4\}$$
$$EF = \{2,4\}$$

$$P(E) + P(F) = \frac{1}{2} + \frac{4}{6} > 1 \qquad\qquad Wrong!$$

$$P(E \cup F) = P(E) + P(F) - P(EF) = \frac{1}{2} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$$

Example

Chevalier de Méré, a writer and nobleman from Louis XIV's court, found that it is a good strategy to bet that a '6' would turn up with one die in 4 throws, but he constantly lost if he betted that two '6' would turn up with two dice in 24 throws. He asked Pascal why? Can you help him?

Hint: You have to use the complement cleverly

$$P(no\ 6) = 1 - \frac{1}{6} = \frac{5}{6}$$

Since the number obtained by a dice in one throw and that obtained in another throw should be "independent"

$$P(no\ 6\ in\ 4\ throws) = \left(\frac{5}{6}\right)^4$$

Using the complement proposition: $P(E^c) = 1 - P(E)$

$$P(at\ least\ one\ 6\ in\ 4\ throws) = 1 - P(no\ 6\ in\ 4\ throws)$$

$$= 1 - \left(\frac{5}{6}\right)^4 = 0.51775 > 0.5$$

Therefore he usually wins

Using two dice, try using the same argument,

$$P(no\ double\ 6\ with\ two\ dice) = 1 - \frac{1}{36} = \frac{35}{36}$$

$$P(no\ double\ 6\ in\ 24\ throws) = \left(\frac{35}{36}\right)^{24}$$

$$P(at\ least\ one\ double\ 6\ in\ 24\ throws) = 1 - \left(\frac{35}{36}\right)^{24} = 0.49140 < 0.5$$

Therefore he usually loses

Sample Spaces Having Equally Likely Outcomes

- For a large number of experiments, it is natural to assume that each point in the sample space is equally likely to occur:

  $S = \{1,2,\ldots,N\}$

  $P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}) = p$

  By Axiom 2 $\quad 1 = P(S)$

  By Axiom 3

  $$P(S) = P(\{1,2,\ldots,N\}) = P(\{1\}) + \cdots + P(\{N\}) = Np$$

  Thus $P(\{i\}) = p = \frac{1}{N}$

  By Axiom 3, for any event E

  $$P(E) = \frac{number\ of\ points\ in\ E}{N}$$

- So we need effective ways to count the number of ways a given event can occur

Basic Principle of Counting

- The mathematical subject of counting is called "combinatorics"
- Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are mn possible outcomes of the two experiments
- The basic principle can be proven by enumerating all the possible outcomes of the two experiments:

$$\begin{array}{l}(1,1),(1,2),\ldots,(1,n)\\(2,1),(2,2),\ldots,(2,n)\\\quad\vdots\\(m,1),(m,2),\ldots,(m,n)\end{array}$$

There are m rows and n columns

Example

Two dice is rolled. The first dice is odd and the second dice is even. How many possible ways of throwing the dice are there?

The first dice may be {1, 3, 5}. The second dice may be {2, 4, 6}. The possible ways are

(1, 2), (1, 4), (1, 6)
(3, 2), (3, 4), (3, 6)
(5, 2), (5, 4), (5, 6)

or 9 possibilities

Generalized Basic Principle of Counting

- If r experiments that are to be performed are such that the first one may result in any of n_1 possible outcomes, and if for each of these n_1 possible outcomes there are n_2 possible outcomes of the second experiment, and if for each of the possible outcomes of the first two experiments there are n_3 possible outcomes of the third experiment, and if, …, then there are a total of $n_1 \cdot n_2 \cdots n_r$ possible outcomes of the r experiments

Example

A password has 8 letters. The first letter must be Capital Letter and the next 4 letters are lower case letters. The last three letters are numbers. How many different passwords are possible?

The first letter may be ‘A’ to ‘Z’. There are 26 possibilities.

Similarly, the second letter may be ‘a’ to ‘z’, which has 26 possibilities.

The numbers may be ‘0’ to ‘9’. There are 10 possibilities.

Thus the number of different passwords are

$$26 \times 26^4 \times 10^3 \approx 1.188 \times 10^{10}$$

Permutation

- Number of different ways n distinct objects can be arranged in a linear order
- Let the n objects be 1, 2, …, n
- In the first position, we have n choices. In the second position, since we have already made a choice in the first position, we have $(n-1)$ choices, …
- There are $n(n-1)\cdots 2\cdot 1$ ordered arrangement
- Each ordered arrangement is called a permutation
- Define $n! = n(n-1)\cdots 2\cdot 1$ $n!$ is called "n factorial"
- $2! = 2\cdot 1 = 2 \quad 1! = 1$
- Define $0! = 1$ (this definition will be useful for later calculations)

Example

Suppose you have five books. In how many ways can the books be put on the bookshelf?

For the first book to be put on the left, we have 5 choices. After you have selected the first book, for the second book, we have four choices, and so on.

Hence there are $(5)(4)(3)(2)(1) = 5!$ ways.

Combination

Suppose now that we are interested in determining the number of different groups of r objects that could be formed from a total of n objects. For instance, how many different groups of three could be selected from the five items A, B, C, D, E? To answer this, reason as follows. Since there are 5 ways to select the initial item, 4 ways to select the next item, and 3 ways to then select the final item, there are thus $5 \cdot 4 \cdot 3$ ways of selecting the group of 3 when the order in which the items are selected is relevant. However, since every group of 3, say the group consisting of items A, B, C, will be counted 6 times (that is, all of the permutations $ABC, ACB, BAC, BCA, CAB,$ CBA will be counted when the order of selection is relevant), it follows that the total number of different groups that can be formed is $(5 \cdot 4 \cdot 3)/(3 \cdot 2 \cdot 1)$

$= 10$

- In general, as $n(n-1)\cdots(n-r+1)$ represents the number of different ways that a group of r items could be selected from n items when the order of selection is considered relevant, and since each group of r items will be counted $r!$ Times in this count, the number of different groups of r items that could be formed from a set of n items is

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}$$

- Define the number of combinations of n objects taken r at a time as $\binom{n}{r} \quad (r \le n)$

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

- Define $\binom{n}{0} = \binom{n}{n} = 1$

Example

There are 120 students in Grade 6 of your secondary school. A Student Committee of 3 students is to be elected from Grade 6 students. How many different possibilities for the composition of the Committee is there?

There are 120 possibilities to choose the first student, 119 possibilities to choose the second student, 118 possibilities to choose the third student. Hence there are (120)(119)(118) possibilities.

However, suppose (Jean, Jane, Janet) are the three students in a committee, then (Jean, Janet, Jane), (Jane, Jean, Janet), (Jane, Janet, Jean), (Janet, Jean, Jane), (Janet, Jane, Jean) are the same committee. Each of the 3! possibilities actually represent the same committee. Hence the number of different committees are

$$\frac{(120)(119)(118)}{3!} = \frac{120!}{3!\,117!} = \binom{120}{3}$$

Example

Suppose 8 students are divided into 2 freely formed groups of 4. How many possible grouping arrangements are there?

The first group has 4 students. There are $\binom{8}{4}$ ways of forming the first group. After the first group is formed, the remaining students form the second group. Hence there are

$$\binom{8}{4} = \frac{8!}{4!\,4!} = 70$$

arrangements

Example

Calculate the winning probabilities for ‘Mark Six’.

Mark Six rules:

49 numbers

1st Prize All 6 drawn numbers

2nd Prize 5 out of 6 drawn numbers + the extra number

3rd Prize 5 out of 6 drawn numbers

⋮

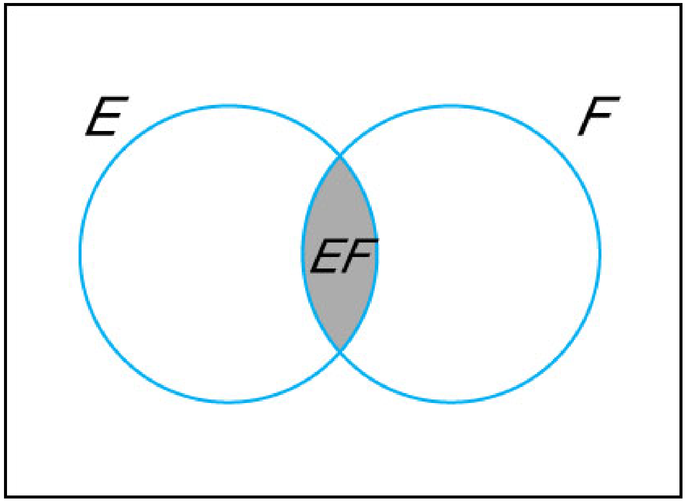
The answer is in https://en.wikipedia.org/wiki/Mark_Six
under ‘Winning Probabilities’ or
Mark Six Probabilities

Conditional Probability

We are often interested in calculating probabilities when some partial information concerning the result of the experiment is available, or in recalculating them in light of additional information

Suppose that one rolls a pair of a fair dice. The sample space S of this experiment is $S = \{(i, j), i = 1,2,3,4,5,6, \;\; j = 1,2,3,4,5,6\}$

Suppose F is the event that the first roll is 3 (condition). Let E be the event that the sum of the two dice rolls is 7. Then there is a 1 in 6 chance that the second roll gives a 4, so that the sum of the two dices is 7. Hence the conditional probability is $\frac{1}{6}$. We denote the conditional probability as $P(E|F)$ (probability of E given that F has occurred)



- If the event F occurs, then in order for E to occur it is necessary that the actual occurrence be a point in both E and F; that is, it must be in EF. However, because we know that F has occurred, it follows that we can regard F as the new sample space and hence the probability that the event EF occurs will equal the probability of EF relative to the probability of F. That is

$$P(E|F) = \frac{P(EF)}{P(F)}$$

- $P(F) > 0$ for the above equation to be well defined

Example

A fair coin is one for which there is a equal probability of turning up “head” and “tail”. If a fair coin is tossed twice, what is the conditional probability that both flips land on “heads”, given that the first flip lands on “heads”?

Let $E = \{(h, h)\}$ be the event that both flips land on “heads”.
Let $F = \{(h, h), (h, t)\}$ be the event that the first flip lands “heads”.

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P\{(h, h)\}}{P\{(h, h), (h, t)\}} = \frac{1}{2}$$

We have to identify the events correctly

Example

A student is taking a 2 hour exam. Suppose the probability that the student will finish the exam in less than x hours is $x/2.5$. Given that the student is still working after 1.5 hours, what is the conditional probability that he can finish the exam within 2 hours?

Let F be the event the student is still working after 1.5 hours. Then F^c is the event that the student finishes in less than 1.5 hours.

$$P(F) = 1 - P(F^c) = 1 - \frac{1.5}{2.5} = 0.4$$

Let E be the event that he finishes the exam within 2 hours. Then EF is the event that he is still working after 1.5 hours but finishes within 2 hours.

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{2}{2.5} - \frac{1.5}{2.5}}{0.4} = 0.5$$

Example

Joseph is undecided whether to take EE4222 Digital Forensics or EE4146 Data Engineering and Learning Systems. He estimates that his probability of getting B+ would be 2/3 in EE4222 and 0.5 in EE4146. If Joseph decides to base his decision on the flip of a fair coin, what is the probability that he gets a B+ in EE4222?

Let F be the event that Joseph takes EE4222 and E be the event that he gets a B+.

$$P(EF) = P(E|F)P(F) = \left(\frac{2}{3}\right)(0.5) = \frac{1}{3}$$

Example

Suppose that a bag contains 4 red balls and 4 white balls. We draw 2 balls from the bag without replacement. If we assume that at each draw each ball is equally likely to be chosen, what is the probability that both balls drawn are white?

Let F be the event that the first ball is white and E be the event that the second ball is white. $P(F) = \frac{4}{8} = 0.5$.

Since the first ball is white, there are 4 red balls and 3 white balls left, hence $P(E|F) = 3/7$

$$P(EF) = P(E|F)P(F) = \left(\frac{3}{7}\right)(0.5) = \frac{3}{14}$$

For more than two events

$$P(E_1 E_2 E_3 \cdots E_n)$$
$$= P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1 \cdots E_{n-1})$$

Proof

Apply the definition of conditional probability to the right hand side

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})}$$

$$= P(E_1E_2E_3\cdots E_n)$$

Example

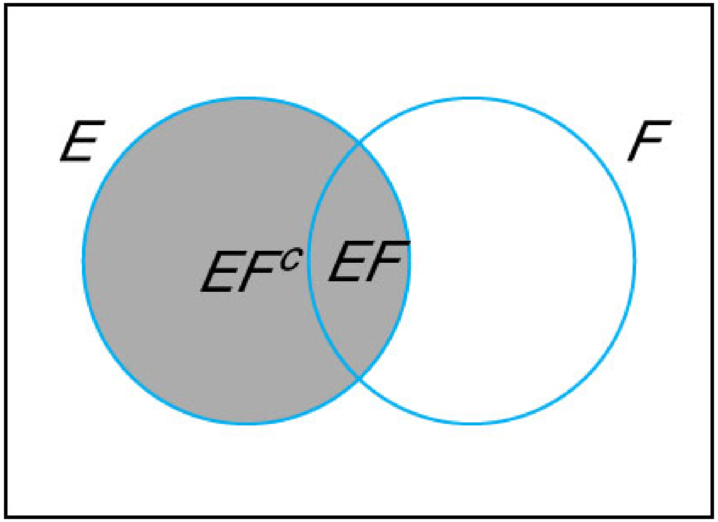
- Estimating the probability of finding ET. Apply the argument to the Drake equation
- The same argument has been applied to estimate the probability of finding a girlfriend

https://www.mathromance.net/topic/drakeequation

https://www.youtube.com/watch?v=ClPPSry8bBw

Bayes' Formula

Discovered by Reverend Thomas Bayes (1702-1761), an English priest interested in philosophical and mathematical problems



Let E and F be events. We may express E as

$$E = EF \cup EF^c$$

For in order for a point to be in E, it must either be in both E and F or be in E but not in F. As EF and EF^c are mutually exclusive, by Axiom 3

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}$$

Physical meaning

- The probability of the event E is a weighted average of the conditional probability of E given that F has occurred $P(E|F)$ and the conditional probability of E given that F has not occurred $P(E|F^c)$, with each conditional probability being given as much weight as the event it is conditioned on has of occurring

- An extremely useful formula, for its use often enables us to determine the probability of an event by first "conditioning" on whether or not some second event has occurred, as the examples below show.

Example

Steve answers a multiple choice question either by knowing the answer or guessing. Suppose the probability that he knows the answer is 0.8 and the probability that he guesses is 0.2. Let there be 4 choices in the question. What is the conditional probability that he knows the answer, given that he answers correctly?

Let E be the event that the Steve knows the answer.
Let F be the event that he answers correctly.

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$P(EF) = P(F|E)P(E)$ =P(FE)=P(E|F)P(F)

P(E) = 0.8 (There is a probability of 0.8 that he knows the answer)

$P(F|E) = 1$ (He will surely answers the question correctly if he knows the answer)

Hence $P(EF) = 0.8$

F can be divided into two mutually exclusive events: He knows the answer or guesses. By applying Bayes' formula

$P(F) = P(F|E)P(E) + P(F|E^c)P(E^c)$

Since there are four choices, if he guesses randomly, $P(F|E^c) = 0.25$

Hence

$$P(F) = (1)(0.8)+(0.25)(0.2)=0.85$$

$$P(E|F) = \frac{0.8}{0.85} \approx 0.94$$

That is, if he answers correctly, we are pretty sure that he knows the answer.

The use of Bayes' reasoning can often help us to correctly access the actual risk, as the next example shows

Example

A laboratory blood test is 95% effective in detecting a certain disease when it is in fact, present. That is, it would give a 5% "false negative" rate. However, the test also yields a "false positive" result for 1% of healthy persons tested. If 0.5% of the population actually has the disease, what is the probability a person has the disease given that the test result is positive?

E F

Let E be the event that the person has the disease.
Let F be the event that the test result is positive.

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)}$$

$$= \frac{(0.95)(0.005)}{(0.95)(0.005) + (0.01)(0.995)} \approx 0.323$$

which is much lower than expected!

Bayesian Inference

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P(EF)}{P(E)}\frac{P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F)}$$

We often know $P(data|hypothesis)$. However, we would like to know $P(hypothesis|data)$. Bayesian inference enables us to do so by

Posterior Probability

Likelihood

Prior Probability

$$P(hypothesis|data) = \frac{P(data|hypothesis)P(hypothesis)}{P(data)}$$

Evidence
(also called marginal likelihood)

If we have a set of hypotheses, and wish to see which hypothesis can best explain the data, we need only to find the hypothesis that maximize $P(hypothesis|data)$. There is no need to compute $P(data)$ since it is constant for every hypothesis

$$P(hypothesis|data) \propto P(data|hypothesis)P(hypothesis)$$

If we have no preference for each hypothesis, then $P(hypothesis)$ is constant

$$P(hypothesis|data) \propto P(data|hypothesis)$$

We find the hypothesis that maximize $P(data|hypothesis)$, i.e., the hypothesis that can best “explain” the data. This is maximum likelihood estimation since it maximizes the likelihood function

Example (a more detailed account can be found in http://jim-stone.staff.shef.ac.uk/BookBayes2012/BayesRuleBookDownloadCh01.html)

Suppose you wake up one day with spots on your face. Suppose $P(spots|smallpox) = 0.9$. Smallpox is a deadly disease, but you are not worried, why?

Smallpox is the first disease to be eradicated from Earth in 1979.

$$P(smallpox|spots) = \frac{P(spots|smallpox)P(smallpox)}{P(spots)}$$

Since $P(smallpox) \to 0$, $P(smallpox|spots) \to 0$

Example

Is this a hill or a crater?

Turn the page upside down. What do you see now?

Explanation: Based on prior experience, we usually think that the light comes from above.

General Form of Bayes' Formula

Bayes' formula can be generalized to more than two events. Let $F_1, F_2, \dots, F_n$ be mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

$$E = \bigcup_{i=1}^{n} EF_i$$

Since EF_i, $i = 1, \dots, n$ are mutually exclusive,

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P\left(E \middle| F_j\right)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

Independent Events

- In the special case where $P(E|F) = P(E)$, E is independent of F. Knowledge that F has occurred does not change the probability that E occurs

$$P(E|F) = \frac{P(EF)}{P(F)} = P(E) \Leftrightarrow P(EF) = P(E)P(F)$$

- Since the equation is symmetric, if E is independent of F, so is F independent of E
- Two events are said to be independent if

$$P(EF) = P(E)P(F)$$

Two events that are not independent are dependent

Example

A card is selected at random from an ordinary deck of 52 playing cards. If E is the event that the selected card is '10' and F is the event that it is a 'heart', show that the two events are independent.

$$P(E) = \frac{4}{52}$$

$$P(F) = \frac{13}{52}$$

$$P(EF) = \frac{1}{52}$$

$$P(EF) = P(E)P(F)$$

Example

Two coins are flipped, and all 4 outcomes are assumed to be equally likely. If E is the event that the first coin lands heads and F is the event that the second lands heads, show that E and F are independent.

$$P(E) = P\{HT, HH\} = \frac{1}{2}$$

$$P(F) = P\{TH, HH\} = \frac{1}{2}$$

$$P(EF) = P\{HH\} = \frac{1}{4} = P(E)P(F)$$

Proposition: If E and F are independent, then so are E and F^c

Proof:

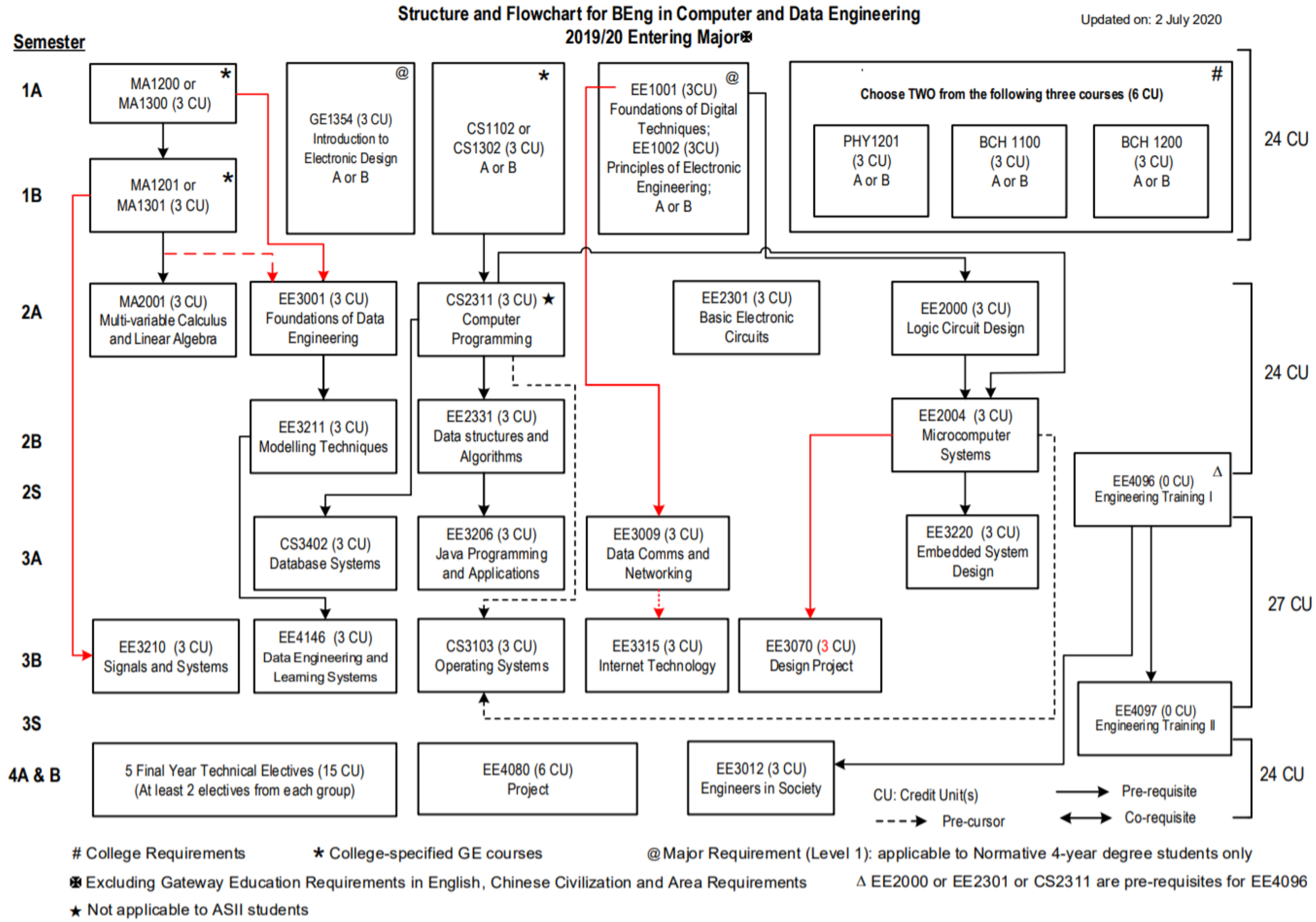
Assume that E and F are independent. Since $E = EF \cup EF^c$, and EF and EF^c are obviously mutually exclusive,

$$P(E) = P(EF) + P(EF^c) = \mathrm{P(E)P(F)} + \mathrm{P(E}F^c)$$

or equivalently,

$$P(EF^c) = P(E)\big(1 - P(F)\big) = P(E)P(F^c)$$

Example



Is the grade of EE2301 and EE3315 independent?

Suppose E is independent of F and is also independent of G. Is E necessarily independent of FG? The answer, somewhat surprisingly, is no.

Example

Two fair dice are thrown. Let

E event that the sum is 7

F event that the first throw is 2

G event that the second throw is 5

Show that E is independent of F, and E is also independent of G, but E is not independent of FG.

E and F are independent because

$$P(E) = P\{(1,6), \dots, (6,1)\} = \frac{6}{36} = \frac{1}{6}$$

$$P(F) = P\{(2,1), \dots, (2,6)\} = \frac{1}{6}$$

$$P(EF) = P\{(2,5)\} = \frac{1}{36} = P(E)P(F)$$

E and G are independent because

$$P(G) = P\{(1,5), \dots, (6,5)\} = \frac{1}{6}$$

$$P(EG) = P\{(2,5)\} = \frac{1}{36} = P(E)P(G)$$

However, $P(E|FG) = 1 \neq P(E)$

Independence involving more than two events

Definition:

Three events E, F, and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

In general, the events $E_1, \ldots, E_n$ are independent if for every subset $E_{1'}, \ldots, E_{r'}, r \le n$, of these events

$$P(E_{1'} \cdots E_{r'}) = \mathrm{P}(E_{1'}) \cdots P(E_{r'})$$

Probability in Daily Life

Challenging the Intuition

- Some probability questions may challenge our intuition. So we need to keep a clear head. Below is one example:

  The Monte Hall Problem: https://www.youtube.com/watch?v=mhlc7peGlGg

- Ch. 2.6 of J.P. Marques De SÁ, Chance, the life of games & the game of life contains more such examples. Please open the E-book in the Course Reserve and read them.

References

- Text book (Available as online book in Course Reserve), Ch. 3
- S.M. Ross, A First course in probability, 9th Edition, Pearson, 2014 (Available as online book in Course Reserve)
- J.P. Marques De SÁ, Chance, the life of games & the game of life, Springer, 2008 (Available as online book in Course Reserve)
- Law of Total Probability and other topics. This youtube author explains well
  https://www.youtube.com/watch?v=7t9jyikrG7w
- A good informal introduction to Bayes' rule is
  http://jim-stone.staff.shef.ac.uk/BookBayes2012/BayesRuleBookDownloadCh01.html