Chapter 1 What is Probability?

- What is probability?
- What is/are statistics?
- Random phenomenaThe origins of the mathematical theory of probability are rooted in the study of games of chance such as dice and card games, in which the results are determined by chance and not by the skill of the player and gambling and began as an attempt to study the random phenomena that have puzzled, intrigued
- Where is all this useful in engineering?
- Deterministic phenomena
  – daily rising of the sun
  – annual flooding of the Nile
  – tides
  – phases of the moon
- Random phenomena
  – tossing coins
  – rolling dice
  – gambling
- Probabilistic notions are commonplace in everyday language
  Probable, improbable, possible, impossible, uncertain, likely, sure, chances of winning are 50-50, odds are 4 to 1, probability of precipitation is 20% for tonight and 40% for tomorrow
- Purely subjective approach to probability: Probabilities are numerical expressions representing our beliefs and biases
- “Probability of precip is 20% for tonight; 40% for tomorrow” = weatherman believes that it is more likely to rain tomorrow than this evening
- What does the following mean? “I am 95% sure that Shakespeare’s plays were written by Francis Bacon”
- Here, either the plays were written by Bacon, or they were not (but I don’t know which!) The 95% probability denotes the strength of my belief in the assertion.
- Purely subjective approach to probability is unsatisfactory in many ways
- Different biases and beliefs, lead to different conclusions from the same data
- However, all approaches to probability are ultimately subjective in some sense
- What is the probability that head will show when I toss this coin? P(Head) = ?
- Most people will answer 1/2
- Why is P(Head) = 1/2?
- There are two common answers to the question: Why 1/2?

1. There are two possible outcomes Head and Tail, so P(Head) = P(Tail) = 1/2
   This is called the *classical approach* to probability (LaPlace)
2. If the coin is tossed repeatedly, Head and Tail will occur (roughly) equally often, so P(Head) = *relative frequency* of Head = 1/2
   This is called the *relative frequency approach* to probability (Poisson, von Mises)

Classical Approach to Probability

- *Experiment* such as tossing a coin or rolling a die
- When the experiment is performed, we observe an *outcome*
  (such as Head, or Tail, or 1, or 2, or 3, or 4, or 5, or 6)
- The set of all possible outcomes that we might observe (on different trials) is the *sample space* of the experiment and is denoted by S or Ω
- **Example:** if the experiment is tossing a coin, then Ω = {Head, Tail}
- If the sample space has n outcomes in it, i.e., $|\Omega| = n$, then each outcome is said to have probability 1/n
- Note that all the n outcomes have the same probability 1/n

- Note that the definition does not tell you what these probabilities *mean*
- LaPlace: the outcomes in Ω are defined to be equally probable because they are equally likely
- The definition is circular! What does “equally likely” mean if not “equally probable”??
- **Subjectivity:** what is an outcome? If the experiment consists of tossing two coins, is Ω = {HH, HT, TH, TT} or is Ω = {0 head, 1 head, 2 heads}
  In the first case, P(both coins show H) = 1/4, while in the second case, P(both coins show H) = 1/3
- **Subjectivity:** what about outcomes such as coin rolled into storm sewer and was never seen again?
- **Subjectivity:**how do we *know* that the outcomes are equally likely? This is a belief!
- Ω = {Win lottery, not win lottery} Equally probable?
- **The classical approach**
- The basic assumptions are valid and often applicable in real-life situations
- Two identical coins are tossed
- Is Ω {HH, HT, TH, TT} or is it {0 head, 1 head, 2 heads}?
- Physical objects that are claimed to be identical in all respects nonetheless behave probabilistically as if they were distinguishable
- Ω = {HH, HT, TH, TT}
- When two dice are rolled, Ω = {(i, j), $1 \le i \le 6$, $1 \le j \le 6$}
- The equally likely assumption is defended on grounds of symmetry or on the basis of the *principle of indifference* or *principle of insufficient reason*
- There is insufficient reason to prefer one outcome over another
- In most instances where symmetry is a valid assumption, the relative frequency approach supports the equally likely outcomes
- Is there more to classical approach than just counting the outcomes and assigning each a probability of 1/n?
- **Event**
- An event A is a collection of outcomes, i.e., a subset of Ω
- Example: If two dice are rolled, the event “rolling a 7” is the subset A = {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)} with six outcomes in it
- On a particular trial of an experiment, an event A is said to have occurred if the outcome observed is a member of A
- **Example:** A 7 is said to have been rolled with two dice if the outcome is ***any*** *one* of {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}
- On an experiment with sample space Ω, we can define lots of different events
- **Example:** If the experiment consists of rolling two dice, the event “3 on the first die” is B = {(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)}
- If the sample space Ω has size n, there are 2^n different possible events
- **Example:** If Ω = {a,b,c}, there are 8 events:
  ø (empty set), {a}, {b}, {c}, {a,b}, {a,c}, {b,c}, {a,b,c} = Ω
- An event A is a collection of outcomes, i.e., a subset of Ω
- On any particular trial of the experiment, the event A is said to have occurred if the outcome is a member of A
- **Example:** Rolling any of (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) with two dice counts as “a 7”
- If |Ω| = n, there are 2^n different possible events
- **Example:** If Ω = {a,b,c}, there are 8 events:
  ø (empty set) {a}, {b}, {c}, {a,b}, {a,c}, {b,c}, {a,b,c} = Ω
- The empty set ø is called the *null event* or the *impossible event* – it never occurs

- The sample space Ω is called the *sure event* or the *certain event* – it always occurs
- Events with one outcome are called *elementary events*
- $|\Omega| = n \leftrightarrow$ n elementary events
- For k = 0, 1, 2, … , n, how many events contain exactly k outcomes?

  $k = 0 \quad \leftrightarrow$ one event ø

  $k = 1 \quad \leftrightarrow$ n elementary events

  $k = n \quad \leftrightarrow$ 1 event Ω
- $\binom{n}{k}$ denotes the number of different events containing exactly k outcomes from the sample space Ω of size n
- $\binom{n}{k}$ is read as "n-choose-k" (number of ways of choosing k things out of n)

  $k = 0 \quad \leftrightarrow$ one event ø

  $k = 1 \quad \leftrightarrow$ n elementary events

  $k = n \quad \leftrightarrow$ 1 event Ω

  $\binom{n}{0} = 1$; $\binom{n}{1} = n$; $\binom{n}{n} = 1$. But, what is $\binom{n}{k}$ in general?
- Start with a subset of size k–1
- Add one of the remaining n–(k–1) elements to the subset to increase size to k
- Any given subset of size k can be obtained in this way from k subsets of size k–1
- Thus, $(n-(k-1)) \times \binom{n}{k-1} = k \times \binom{n}{k}$
- $\binom{n}{k} = \binom{n}{k-1} \times \frac{n-(k-1)}{k} = \frac{n}{1} \times \frac{n-1}{2} \times \frac{n-2}{3} \times \cdots \times \frac{n-k+1}{k} = \frac{n}{1} \times \frac{n-1}{2} \times \cdots \times \frac{n-k+1}{k} \times \frac{n-k}{n-k} \times \cdots \times \frac{1}{1}$

  $= \frac{n!}{k!(n-k)!}$ is also known as a *binomial coefficient*
- **Binomial theorem:** $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$
- Putting x = y = 1, we get $2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n}$

  Left side = total # of events

  Right side counts the total # of events in order of size
- **Complementary event:** If A is an event, the set of all outcomes that are not in A is called the complement of A
- $A^c = \{x: x \quad \Omega, x \quad A\}$
- If $|A| = k$, then $|A^c| = n-k$
- $\binom{n}{k} = \binom{n}{n-k}$; Remember $\frac{n!}{k!(n-k)!}$
- The complement of A^c is A. $(A^c)^c = A$
- If A occurs, then A^c does not. If A^c occurs, then A does not.
- x A if and only if x A^c
- For any given event A, on a trial of the experiment, one of A and A^c occurs while the other does not
- On any trial of the experiment exactly half of the 2^n events occur and the other half do not

- **Example:** $\Omega = \{a,b,c\}$. If the outcome is b, the 4 events {b}, {b,c}, {a,b}, and {a,b,c} occur while the 4 events ø {a}, {c}, and {a,c} do not
- The probability of an event A is defined to be the size of A divided by the size of Ω
- $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n} = 1 - P(A^c)$
- P(ø) = 0
- $P(\Omega) = 1$
- P(elementary event) = P(outcome) = 1/n
- **Example:** The experiment consists of rolling two dice. There are 36 outcomes.
  $\Omega = \{(i, j), 1 \le i \le 6, 1 \le j \le 6\}$
- P("roll a 7") = 6/36 = 1/6 because the event "roll a 7" has 6 outcomes (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) in it

  A = "roll a 7" = {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}
  B = "3 on first die" = {(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)}
  C = "same on both dice" = {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}
- P(A) = P(B) = P(C) = 6/36 = 1/6
- P("roll a 7 or same on both dice") = P(A   C) = 12/36 = 1/3 since A   C has 12 outcomes

  Note that P(A   C) = P(A) + P(C)
- P("roll a 7 or 3 on first die") = P(A   B) = 11/36 since A   B has only 11 outcomes in it: {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1), (3,1), (3,2), (3,3), (3,5), (3,6)}

  P(A   B) = 11/36 ≠ P(A) + P(B). Instead, P(A   B) = P(A) + P(B) – $P(A \cap B)$
- If two events have no elements in common, that is, their intersection is the empty set, then the events are said to be *disjoint* or *mutually exclusive*
- Events A = {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)} and
  C = {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)} are *mutually exclusive* events
- In general, if events A and B are *mutually exclusive*, then

  P(A   B) = P(A) + P(B)
- But if events A and B are ***not*** mutually exclusive, then

  P(A   B) = P(A) + P(B) – $P(A \cap B)$
- If A and B are mutually exclusive events, this reduces to the previous formula because $P(A \cap B)$ = P(ø) = 0
- **Example:** An urn contains 6 red balls and 4 white balls. A ball is drawn at random. What is the probability that it is red?

  Comment: "random" generally means that the classical approach is to be used and all outcomes are equally likely

  There are ten different outcomes of this experiment (not two!) and
  Ω = {R1, R2, R3, R4, R5, R6, W1, W2, W3, W4} *and not* Ω = {R, W}
  P(Red Ball) = 6/10
- What if the experiment consists of drawing two balls at random?
- Important question to be decided: How are these balls drawn? Is the first one drawn put back in the urn before the second is picked or not? Or are we reaching in and grabbing two at once?
- **Sampling with replacement:** first ball is replaced in urn before the second ball is drawn

  Ω has 100 elements that are *vectors* of the form (X,Y) with X and Y taking on all possible values from the set

  {R1, R2, R3, R4, R5, R6, W1, W2, W3, W4}

  If (X,Y) = (R1,W3), this means the first ball drawn was R1 and the second ball drawn was W3

It is possible for (X,Y) to be (say) the vector (R3,R3) indicating that the same red ball R3 was drawn the second time as well

P(both balls drawn are red) = 6×6/100 = 9/25

P(both balls drawn are white) = 4×4/100 = 4/25

P(one red and one white) = 1 – 9/25 – 4/25 = 12/25

More directly, we get the answer as 6×4/100 + 4×6/100 = 48/100 = 12/25

P(first ball is red) = 6/10

- **Sampling without replacement**: If the first ball drawn is not replaced before the second is drawn, the sample space has 10×9 = 90 elements that are *vectors* of the form (X,Y) with X and Y taking on all possible *different* values from the set

  {R1, R2, R3, R4, R5, R6, W1, W2, W3, W4}

  If (X,Y) = (R1,W3), this means the first ball drawn was R1 and the second ball drawn was W3

  Note that it is *not* possible for (X,Y) to be the vector (R3,R3) because R3 is not being put back into the urn before the second ball is drawn

  P(both balls drawn are red) = 6×5/90 = 1/3

  P(both balls drawn are white) = 4×3/90 = 2/15

  P(one red and one white) = 1 – 1/3 – 2/15 = 8/15

  More directly, we get the answer as 6×4/90 + 4×6/90 = 48/90 = 8/15

  P(first ball is red) = 6/10
- If we reach in and grab two at the same time, we get a (randomly chosen) subset of size 2 from a set of size 10
- Now, each element of Ω is *a set* of the form {X,Y} with X and Y taking on all possible *distinct* values from the set

  {R1, R2, R3, R4, R5, R6, W1, W2, W3, W4}

  Note: Previously, the elements of Ω were vectors (X,Y) where X was the first ball drawn and Y the second, and (R3,W1) was a different outcome from (W1,R3). Now, the elements of Ω are *sets* {X,Y} and we no longer can talk about the first ball drawn and the second ball drawn; both balls are drawn at the same time, and {R3,W1} is the same set as {W1,R3}

  Ω has $\binom{10}{2}$ = 45 elements

  P(both balls drawn are red) = $\binom{6}{2} \div \binom{10}{2} = \dfrac{6\times 5}{10\times 9}$ = 1/3

  P(both balls drawn are white) = $\binom{4}{2} \div \binom{10}{2} = \dfrac{4\times 3}{10\times 9}$ = 2/15

  P(one red and one white) = 1 – 1/3 – 2/15 = 8/15

  P(first ball is red) makes no sense here
- The sample space for the experiment consisting of grabbing random subsets is smaller than the sample space for the experiment consisting of sampling without replacement by a factor of k! where k is the number of items drawn. Essentially, we lose the sense of the order in which items are drawn but the probabilities which can be defined for both experiments are the same in each case. Which way the sample space is defined is often a matter of preference
- For beginning students, the most difficult part is in defining the sample space correctly and in counting the number of elements in various events
- There is a vast literature on this subject. Work through all the examples in Chapter 2.5 of the text

- **Problems with the classical approach**
- All outcomes must have the same probability
- Modification: The n outcomes have probabilities p_1, p_2, … p_n where each $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$
- P(A) = Sum of the probabilities of the outcomes in A
- It is still true that $P(ø) = 0$ and $P(\Omega) = 1$
- It is still true that $P(A \quad B) = P(A) + P(B) - P(A \cap B)$ and $P(A^c) = 1 - P(A)$
- It is impossible to extend the classical approach to sample spaces with infinitely many outcomes – in particular, when the outcomes are real numbers
- Knowing only the classical approach is not very useful in engineering applications

The Relative Frequency Approach

- Experiment with sample space Ω
- Events are subsets of Ω
- N *independent* trials of the experiment are carried out
- **Independence:** the outcome of any trial does not affect the outcome of any other
- **Example:** Sampling *with* replacement gives rise to independent trials; sampling without replacement does not
- If event A is observed to have occurred N_A times during the N trials, its *relative frequency* is $\frac{N_A}{N}$
- $0 \leq N_A \leq N \qquad 0 \leq \frac{N_A}{N} \leq 1$ • $N_ø = 0 \qquad \frac{N_ø}{N} = 0$ • $N_\Omega = N \qquad \frac{N_\Omega}{N} = 1$
- If A and B are *mutually exclusive* events that occur N_A and N_B times respectively on N trials, then A   B occurs $N_A + N_B$ times on the N trials, that is, $N_{A \quad B} = N_A + N_B$ and $\frac{N_{A \quad B}}{N} = \frac{N_A}{N} + \frac{N_B}{N}$
- Relative frequencies behave pretty much like classical probabilities: each relative frequency is between 0 and 1, ø and Ω have relative frequencies 0 and 1 respectively, and the relative frequency of mutually exclusive events is the sum of the relative frequencies
- Define the probability of an event A as its measured relative frequency on N trials
- **Philosophical questions:** Do probabilities exist only in terms of relative frequencies? Can we talk of the probability of Heads for a brand-new never-tossed-before coin?
- What should N be?
- If we toss a coin twice and it turns up heads both times, should we set P(Heads) = 1? Is it reasonable to set P(Heads) = 1 for the 3rd trial, and change it for the 4th?
- Everybody agrees that N should be "large"
- If we toss a coin 100,000 times and it turns up heads 50,025 times, should we set P(Heads) = 0.50025? or should we "fudge the data" and say P(Heads) = 0.5 exactly?
- **Practical problems:** How large should N be? How exact should the answer be?
- Empirical fact: If independent trials of an experiment are carried out, the relative frequency of an event A converges quite rapidly to a constant value
- Relative frequency approach: Define P(A) = constant value
- Relative frequency approach: $P(A) = \lim_{N \to \infty} \frac{N_A}{N}$
- This looks very mathematical but it *cannot* be proved that the limit exists in all cases
- In practice, we can only repeat the experiment a finite number of times

- Subjectivity creeps in again: That relative frequencies will always converge is a belief; convergence cannot be proved (in the mathematical sense)

  That we can deduce what the "limit" is based on only a finite number of observations is a belief

  That the results of our N trials are representative of what will happen in the future is a belief
- Any arbitrary "capricious" finite sequence of results can be compensated for in the infinite (not as yet observed) sequence of outcomes
- A coin that turns up heads a 1,000,000 times in succession is *not necessarily* a biased coin – the 1,000,000 heads could be followed by 500,000,000 tails and 499,000,000 heads, so that after 1,000,000,000 trials, the relative frequency of heads would be 0.5 exactly
- Counterargument: A fair coin cannot possibly turn up heads *that* many times in succession
- Countercounterargument: What is the maximum number of successive heads you are willing to accept? How does the coin know this limitation?

  The *definition* of probability as a relative frequency also runs into trouble when we consider sample spaces with infinitely many outcomes
- **Example:** A coin is tossed repeatedly until Heads shows (for the first time)

  $\Omega = \{H, TH, TTH, TTTH, \ldots\}$
- In a *finite* number of trials, we will observe only a finite number of outcomes. If probability is defined as a relative frequency, then we either cannot assign probabilities to the outcomes that did not occur, or must assign them subjectively
- For the reasons discussed above, the *definition* of probability purely in terms of relative frequencies is not commonly used these days
- However, the notion of relative frequency is an important one, and is used to relate the mathematical theory of probability to reality
- In practice, the notion of relative frequency is used to measure probabilities: the *gathering* of the raw data and the *inferring* of probabilities from these data is the domain of *statistics*
- Practical problems: How large should N be? How exact should the answer be?
- Neither the classical approach of equally likely outcomes nor the relative frequency approach are quite satisfactory as a *definition* for probability
- How should we define probability? We don't define it at all!!

  "I may not know how to define it exactly, but I know it when I see it"

  Justice Potter Stewart, U.S. Supreme Court
- In the axiomatic theory, probabilities are numbers in the range [0,1] and no other *meaning* is assigned to them
- The axiomatic approach does not tell you what probability means, but for any personal view of the meaning of probability that you may have, it provides you with a consistent, rational, and mathematically correct approach to calculation and problem solving
- In the axiomatic theory, we take some probabilities as being given (possibly these might be measured via relative frequency methods, or they may be subjective probabilities reflecting our biases and beliefs) and we use these to *calculate* other probabilities of interest
- Why *calculate* instead of measure probabilities?
- **Example:** it might be easier and cheaper to *calculate* the reliability of a complex system from the reliabilities of its simple components than to *measure* the reliability via N copies of the complex system
- The axiomatic theory of probability builds a suitable model for the rational study of random phenomena
- The mathematical models that have been devised do not include *all* aspects of the phenomena

V = IR

- Several physical properties of actual resistors are not included in this simple Ohm's Law model
- Results obtained from *any* model *do not prove* anything. Results should be interpreted and used with care
- The axiomatic theory cannot provide a *definitive* answer to the question: Is this a fair coin? It can merely provide strong grounds for belief one way or the other

The axiomatic approach to probability

- Experiment with sample space Ω
- Events are subsets of Ω
- Probabilities are defined for all the events subject to the *axioms* of probability theory
- The axioms are just the various properties that we have agreed that probabilities have
- **Axiom I:** For all events A, $0 \leq P(A) \leq 1$

  This is the most fundamental fact about probabilities and you *must* know it and bear it in mind at all times
- **Axiom II:** $P(\Omega) = 1$
- **Axiom III:** If $A_1, A_2, A_3, \ldots, A_n, \ldots$ is a *countable* sequence of *mutually exclusive* events, then

  $P(A_1 \quad A_2 \quad A_3 \quad \ldots \quad A_n \quad \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) + \ldots$
- As a special case of Axiom III we shall show that if A and B are *mutually exclusive* events, then $P(A \quad B) = P(A) + P(B)$
- What have we gained by using axioms instead of just winging it?
- The axioms assert that the same rules apply regardless of *how* the assignments of probabilities were made initially
- The axioms provide for means of dealing with sample spaces with infinitely many outcomes
- Events are subsets of Ω
- An event is said to have occurred if the outcome of the experiment that occurred belongs to the event
- The axiomatic approach does not *require* that all subsets of Ω be events

  Previously, we took all 2^n subsets of Ω to be events, but we don't *have* to
- If certain subsets of Ω are non-events, then it is not necessary to define a probability for these subsets

  Axiom I says: For all *events* A, $0 \leq P(A) \leq 1$

  so if A is not an event, we don't need to specify a probability for it
- Are we allowed to pick and choose which subsets of Ω we call events and which we don't arbitrarily, or are there restrictions on our choice?
- Yes, there are restrictions

1. Ω is always an event

   Otherwise Axiom II: $P(\Omega) = 1$ does not make much sense
2. If we choose to call a subset A an event, then we must also call the complementary subset A^c an event

   If we can talk about A occurring, we also can talk about A not occurring
3. If the *countable* collection of subsets $A_1, A_2, A_3, \ldots, A_n, \ldots$ are all called events, then

   $A_1 \quad A_2 \quad A_3 \quad \ldots \quad A_n \quad \ldots$ and $A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n \cap \ldots$ must also be called events

   At least one of the A_i's occurring, and *all* of the A_i's occurring are also events

- Any collection of events satisfying these three rules is called a σ-algebra of events, or a σ-field of events, or a family of events
- The σ-field contains Ω and is closed under the operations of *complementation* and countable *unions* and *intersections*
- If the sample space Ω is a finite set, then the collection of 2^n subsets of Ω is a σ-field
- However, a smaller collection of subsets can do just as well
- Example: If A is a subset of Ω, then the collection of *four* sets {ø, A, A^c, Ω} is a σ-field
- Example: If A and B are subsets of Ω, then the collection of 16 = 2^4 different unions of *4 minsets* $A \cap B$, $A^c \cap B$, $A \cap B^c$ and $A^c \cap B^c$ is a σ-field
- Compare minterms in a Boolean algebra xy, $\overline{x}y$, $x\overline{y}$, $\overline{x}\,\overline{y}$
- More generally, if we choose to call A, B, C, etc as events, then their complements and countable unions and intersections must also be called events and included in the σ-field, but we are *not required* to include any other subsets in the σ-field
- If the sample space is an interval of real numbers, for example, Ω = (0,1) or Ω = (–∞,∞), then we *cannot* take the σ-field to be the collection of all possible subsets of Ω
- Why not? Because the theory becomes logically inconsistent
- Different perfectly correct ways of calculating the probability of an event can give different results
- What should we choose as the events if Ω is an interval of real numbers?
- Consider a large number of trials of an experiment with sample space Ω = (0,1); for example, 10,000 successive outputs from a so-called "random number generator"
- What is the relative frequency of 0.7071?
- Either 0.7071 will have occurred *once* in the 10,000 trials or it will have *not occurred at all*
- Relative frequency = 10^{-4} or 0
- Observe the next 90,000 outputs
- Relative frequency = 10^{-5} or 0
- Observe the next 900,000 outputs
- Relative frequency = 10^{-6} or 0
- It would appear that the relative frequency of the outcome 0.7071 is converging to a limit of 0 (or is there already!)
- The relative frequency of the outcome 0.68321956 is *also* converging to a limit of 0 (or is there already!) as is the relative frequency of 0.21732
- If the sample space is an interval of real numbers, *every* outcome seems to have probability 0 in the relative frequency sense
- If *all* the outcomes in Ω have probability 0, where did the probability disappear to?
- *Subintervals* of Ω have nonzero probabilities
- Roughly 50% of the observed output of the random number generator is in the interval (0,0.5), 40% in (0.3,0.7), etc
- When Ω is an interval of real numbers, the σ-field is defined to include *all* subintervals, i.e. subsets of the form (a,b) or [a,b) or (a,b] or [a,b], and all other subsets of Ω that must be included on account of Rules 1, 2, amd 3

  Singleton subsets {x} can be expressed as the interval [x,x], and are therefore events

- Are there sets of real numbers that cannot be expressed as countable unions and intersections of intervals?
- Yes: these (weird) subsets are precisely the ones that make the theory inconsistent
- By not including such sets as events, we avoid having to define their probabilities, and the inconsistencies disappear
- Curious fact: The probability of observing 0.7071 is 0 and yet we *could* observe this outcome
- Moral: Events of probability 0 are not impossible events — they will either never occur at all or if they do occur once, they will never occur again
- Since every interval is an event, we can talk of the probability that the outcome x satisfies $a < x < b$, that is, *lies in a certain range*
- These are the *only* questions that make engineering sense
- P(outcome = 0.7071 *exactly)* is a useless notion; besides the answer is 0 anyway!
- A meter reading of 0.7071 volts usually means that the voltage is *between* 0.70705 and 0.70715 volts, and *not* that the voltage is *exactly* 0.7071 volts
- All physical measurements necessarily give discrete rational values of the measured parameter
- All computer-generated numbers (even floating-point numbers) are necessarily discrete rational numbers
- Some parameters are indeed discrete-valued (e.g. current)
- However, it is convenient to model the parameters as continuous variables so that the methods of calculus can be applied
- Thus, $L\frac{di}{dt} + Ri = E$
- We follow the same path and use continuous models for parameters that we measure only in discrete quanta
- We agree to think only of questions such as the probability that $a < x < b$ and not of the probability that $x = c$ exactly

Summary of the axiomatic approach

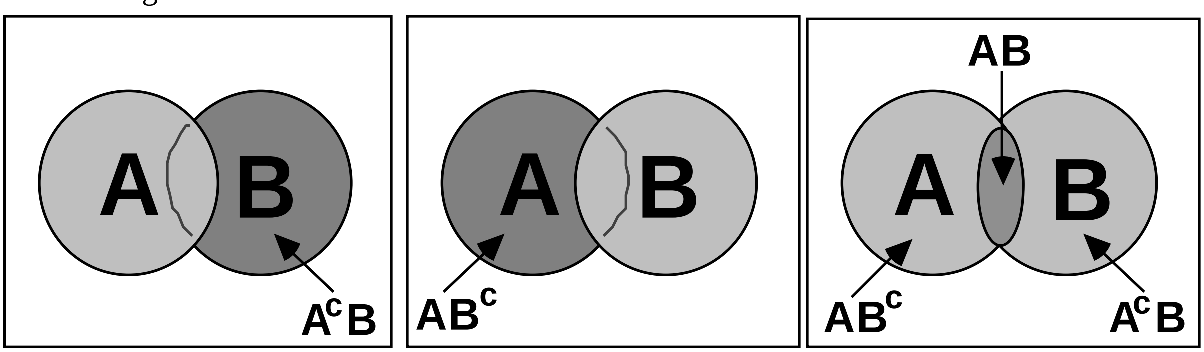
- The axiomatic approach to probability considers three things that together constitute the *Probability space* (Ω, F, P)
- *Sample space* Ω consisting of all possible outcomes of the experiment
- Events are subsets of Ω
- An event is said to have occurred if the outcome of the experiment belongs to the event
- Not every subset of Ω is required to be an event
- *The* σ*-field F* (collection of events defined on Ω) always includes Ω, and is closed under complementation and under countable unions and intersections
- If Ω is an interval of real numbers, the σ-field includes all subintervals and their countable unions and intersections
- The *probability measure* P(•) is a function that assigns real numbers (probabilities) to the events in *F* subject to the *axioms of probability theory*
- **Axiom I:** For all events A $\quad F$, $0 \le P(A) \le 1$
- **Axiom II:** $P(\Omega) = 1$
- **Axiom III:** If $A_1, A_2, A_3, \ldots, A_n, \ldots \quad F$ is a *countable* sequence of *mutually exclusive* events, that is, $A_i \cap A_j = \emptyset$ for $i \neq j$, then

  $P(A_1 \quad A_2 \quad A_3 \quad \ldots \quad A_n \quad \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) + \ldots$

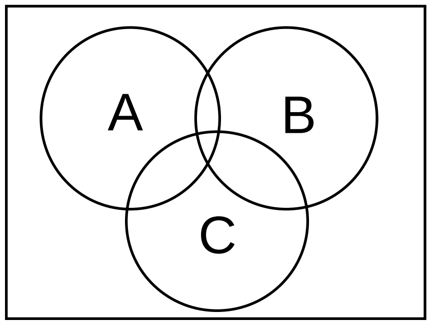
Consequences of the axioms

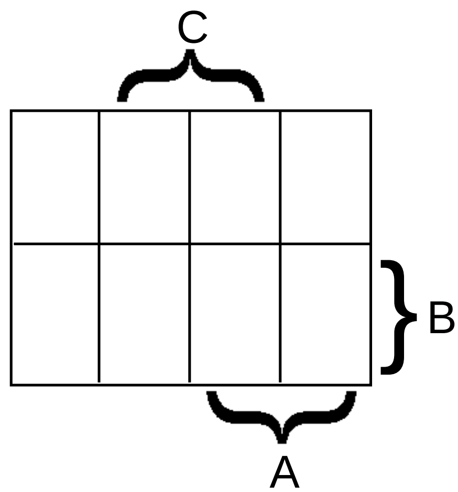
- Several consequences of the axioms provide results which are very useful in doing calculations involving probabilities
- We have already encountered these before
- You should read the proofs (Ross, pp. 32-36) for yourself
- $P(\emptyset) = 0$
- If A and B are mutually exclusive events, then $P(A \quad B) = P(A) + P(B)$
- More generally, the probability of the union of n mutually exclusive events is the sum of their probabilities
- $P(A^c) = 1 - P(A)$
- If $A \quad B$, then $P(A) \le P(B)$
- $P(A \quad B) = P(A) + P(B) - P(A \cap B)$
- The principle of inclusion and exclusion generalizes this result to the union of n sets (Ross, Proposition 4.4, p. 37)
- Read this result and its proof for yourself
- **Useful set-theoretic results:**

  DeMorgan's Laws: $(A \quad B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \quad B^c$
- $A \quad B = A \quad (A^c \cap B)$
- This allows us to break up a set $A \quad B$ into two mutually exclusive subsets
- $A \quad B = A \quad (A^c \cap B) = B \quad (A \cap B^c) = (A \cap B^c) \quad (A^c \cap B) \quad (A \cap B) = AB^c \quad A^cB \quad AB$
- Venn diagrams



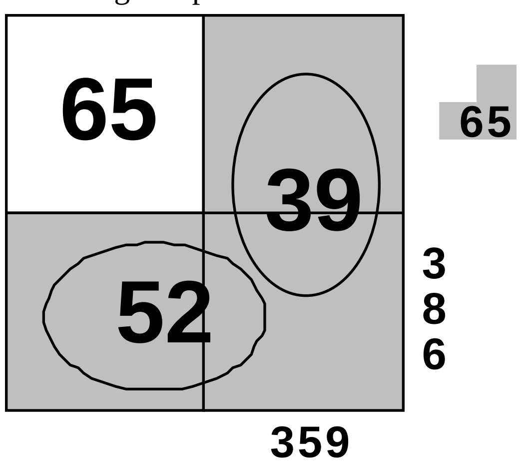
- *Partition* of a set: a collection of mutually exclusive subsets whose union is the set
- AB and AB^c partition A
- ABC, ABC^c, AB^cC, AB^cC^c is a (finer) partition of A
- ABC, ABC^c, AB^cC, AB^cC^c, A^cBC, … , $A^cB^cC^c$ partition Ω
- Use Karnaugh maps instead of Venn diagrams as shown on the next page

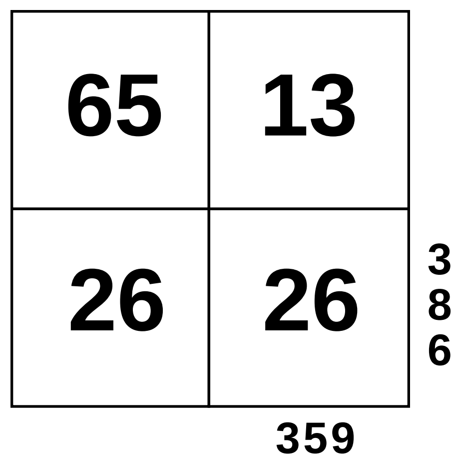




- **Puzzle:** 39 of the 130 students in ECE 350 are also taking ECE 359, 52 of them are also taking ECE 386, and 65 are taking neither ECE 359 nor ECE 386. How many students are taking all three courses?
- **Solution:** Since 65 of the 130 students in ECE 350 are in neither ECE 359 nor ECE 386, the remaining 65 must be in either ECE 359 or ECE 386 or both.

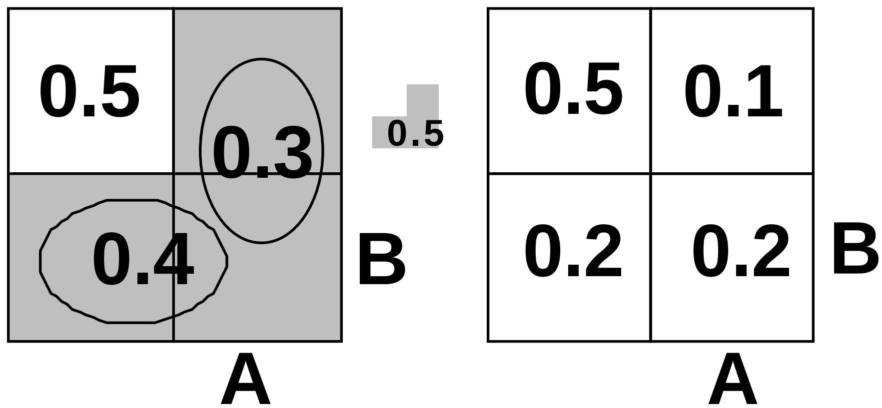
  The students in both ECE 359 and ECE 386 are counted among the 39 students in ECE 359 as well as among the 52 in ECE 386. Hence, 26 students must be in both ECE 386 and ECE 359; another 26 must be in ECE 386 but not in ECE 359 (giving 26+26 = 52 in ECE 386); another 13 must be in ECE 359 but not in ECE 386 (giving 26+13 = 39 in ECE 359)
- A Karnaugh map solution is as follows:





- **Example 1:** A student is picked at random from among the 130 students in ECE 350. Let A = “student is in ECE 359”, B = “ student is in ECE 386”. If P(A) = 0.3 and P(B) = 0.4, and $P(A^cB^c) = 0.5$, i.e., $|A| = 39$, $|B| = 52$ and $|A^cB^c| = 65$, what is P(AB)?
- **Solution:** $A^cB^c = (A \quad B)^c$ and hence $P(A \quad B) = 1 – P((A \quad B)^c) = 1 – P(A^cB^c) = 0.5$

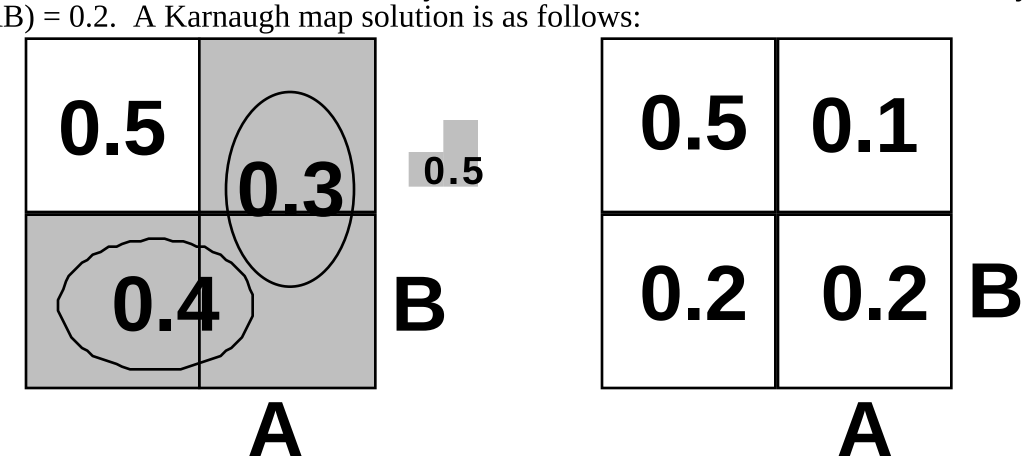
  $P(A \quad B) = P(A) + P(B) – P(AB)$ and hence $P(AB) = 0.3 + 0.4 – 0.5 = 0.2$ showing that there are 26 students in the set AB
- A Karnaugh map solution is as follows:



- **Example 2:** Let A = “student is in ECE 359” B = “ student is in ECE 386”. If P(A) = 0.3, P(B) = 0.4, and $P(A^cB^c) = 0.5$, what is P(AB)?
- The problem does not *specify* how many students there are (i.e. what $|\Omega|$ is) or whether the students are equally likely to be picked. No details of the

  σ-field are given (except that A and B obviously are events of interest)
- **Solution:** $A^cB^c = (A \quad B)^c$ and hence $P(A \quad B) = 1 – P((A \quad B)^c) = 1 – P(A^cB^c) = 0.5$.

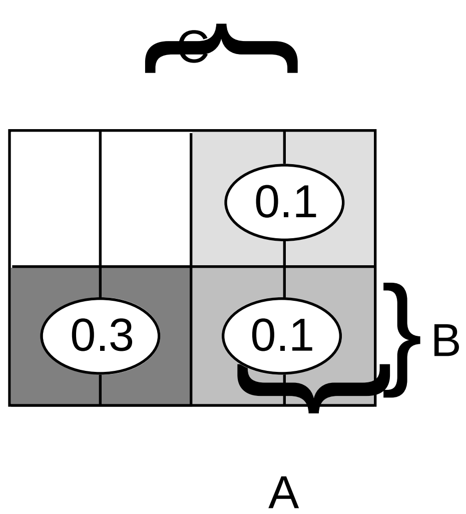
  $P(A \quad B) = P(A) + P(B) – P(AB)$ and hence $P(AB) = 0.3 + 0.4 – 0.5 = 0.2$
  Notice that we do not know how many students there are in the set AB; all we can say is that P(AB) = 0.2. A Karnaugh map solution is as follows:

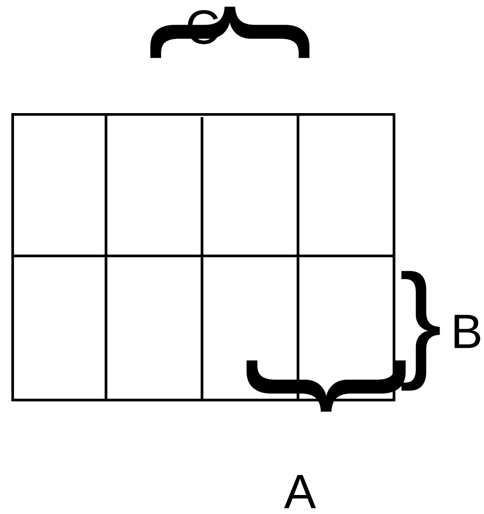


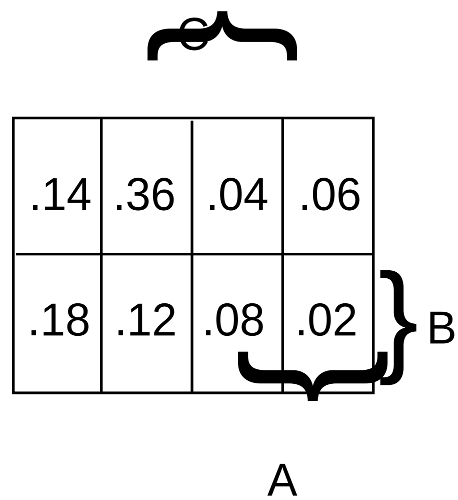
- **Example 3:** Events A and B are defined on a sample space Ω. If P(A) = 0.3, P(B) = 0.4, and $P(A^cB^c) = 0.5$, what is P(AB)?
- **Solution:** Exactly the same as before! However, now we do not even have any storybook “meaning” to assign to the event AB

- **Example 4:** Events A, B, C are defined on a sample space Ω. Given P(A) = 0.2, P(B) = 0.4, P(C) = 0.6, P(AB) = 0.1, P(B   C) = 0.8, P(A   C) = 0.68, and P(ABC) = 0.08, what is the probability that event A occurs, but neither of the events B and C occurs? What is the probability that at least two of the three occur? What is the probability that none of the three events occurs?

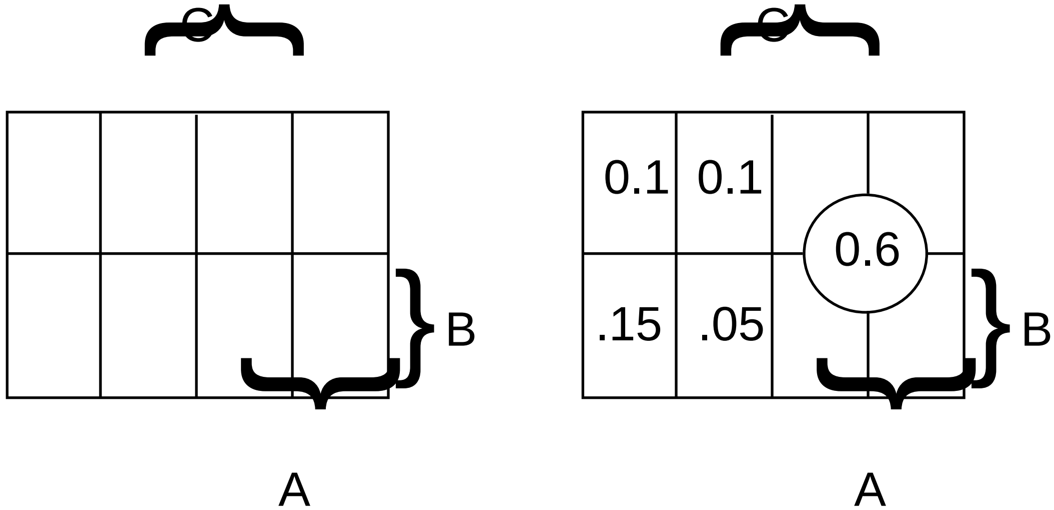
  We are asked for the probabilities of the events AB^cC^c, AB   BC   AC, $A^cB^cC^c$



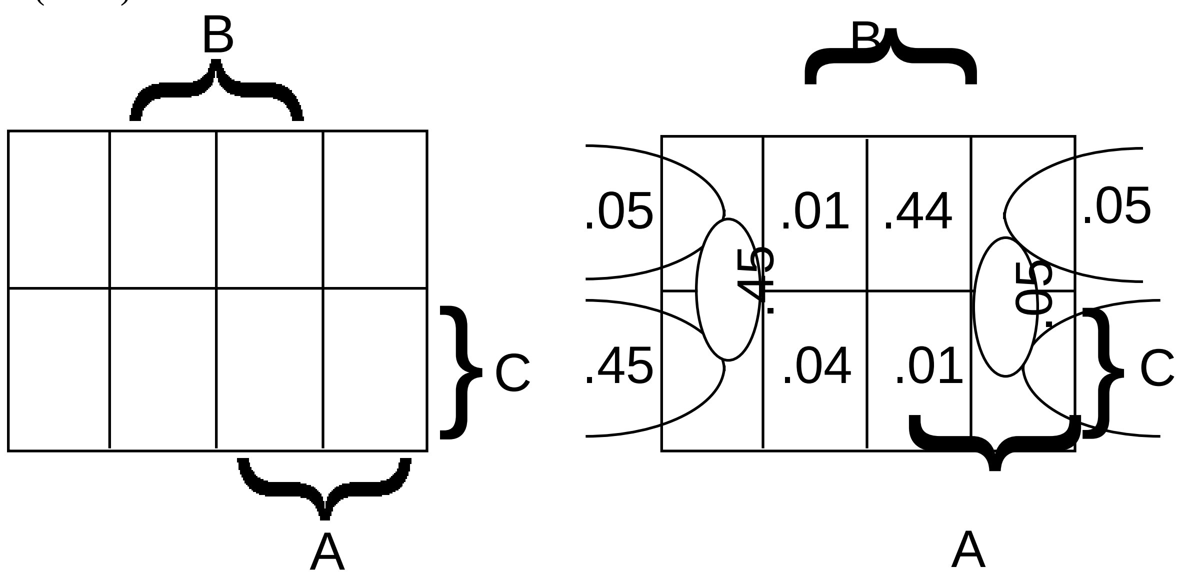




- **Example 5:** Events A, B, C, are defined on a sample space. If $P(A) = 0.6$, $P(A \ \ B^c) = 0.8$, $P(A \ \ B \ \ C) = 0.9$, and $P(A^cC) = 0.15$, find $P(A^cBC)$



- **Example 6:** A, B, C are events defined on a sample space. If $P(A) = P(B) = P(C) = 0.5$, $P(B^cC^c) = 0.05$, $P(ABC) = 0.01$, and $P((A \ \ B^c \ \ C)^c) = 0.01$, find $P(AB)$, $P(AB^c)$, and $P(A^cBC)$.



- **Example 7:** Now suppose that A, B, and C denote the events that the inputs and the output of a (possibly malfunctioning) 2-input NAND gate are 1. If the event $A^cB^cC^c$ occurs, is the NAND gate functioning correctly, i.e., is the output value the NAND of the input values ? Let E = event that NAND gate is functioning incorrectly. Compute P(E).