Motivation and Applications: Why Should I Study Probability?

- As stated by Laplace, “Probability is common sense reduced to calculation”.
- You need to first learn the theory required to correctly do these calculations. The examples that I solve and those in the book and the homeworks will provide a wonderful practical motivation as to why you need to learn the theory.
- If you patiently grasp the basics, especially the first 4 chapters of BT, it will be the most useful thing you’ve ever learnt - whether you pursue a career in EE or CE or Economics or Finance or Management and also while you try to invest in stocks or gamble in Las Vegas!
- Applications: communications (telephones, cell phones, TV, ...), signal processing (image and video denoising, face recognition, tracking

moving objects from video,...), systems and control (operating an airplane, fault detection in a system,...), predicting reliability of a system (e.g. electrical system), resource allocation, internet protocols, non-engineering applications (e.g. medicine: predicting how prevalent a disease is or well a drug works, weather forecasting, economics).

Introduction: Topics Covered. Chapter 1, 1.1 - 1.6)

- What is Probability
- Set Theory Basics
- Probability Models
- Conditional Probability
- Total Probability and Bayes Rule
- Independence
- Counting

What is Probability?

- Measured relative frequency of occurrence of an event.
  Example: toss a coin 100 times, measure frequency of heads or compute probability of raining on a particular day and month (using past years' data)
- Or subjective belief about how "likely" an event is (when do not have data to estimate frequency).
  Example: any one-time event in history or "how likely is it that a new experimental drug will work?"
  This may either be a subjective belief or derived from the physics, for e.g. if I flip a symmetric coin (equal weight on both sides), I will get a head with probability $1/2$.
- For probabilistic reasoning, **two** types of problems need to be solved

1. Specify the probability “model” or learn it (covered in a statistics class).
2. Use the “model” to compute probability of different events (covered here).

- We will assume the model is given and will focus on problem 2. in this course.

Set Theory Basics

- Set: any collection of objects (elements of a set).
- Discrete sets
  - Finite number of elements, e.g. numbers of a die
  - Or infinite but countable number of elements, e.g. set of integers
- Continuous sets
  - Cannot count the number of elements, e.g. all real numbers between 0 and 1.
- "Universe" (denoted Ω): consists of all possible elements that could be of interest. In case of random experiments, it is the set of all possible outcomes. Example: for coin tosses, $\Omega = \{H, T\}$.
- Empty set (denoted ϕ): a set with no elements

- Subset: $A \subseteq B$: if every element of A also belongs to B.
- Strict subset: $A \subset B$: if every element of A also belongs to B and B has more elements than A.
- Belongs: $\in$, Does not belong: $\notin$
- Complement: A' or A^c, Union: $A \cup B$, Intersection: $A \cap B$
  - $A' \triangleq \{x \in \Omega | x \notin A\}$
  - $A \cup B \triangleq \{x | x \in A, \ or \ x \in B\}$, $x \in \Omega$ is assumed.
  - $A \cap B \triangleq \{x | x \in A, \ and \ x \in B\}$
  - Visualize using Venn diagrams (see book)
- **Disjoint sets: A and B are disjoint if $A \cap B = \phi$ (empty), i.e. they have no common elements.**

- DeMorgan's Laws

$$(A \cup B)' = A' \cap B' \quad (1)$$
$$(A \cap B)' = A' \cup B' \quad (2)$$

  - Proofs: Need to show that every element of LHS (left hand side) is also an element of RHS (right hand side), i.e. LHS $\subseteq$ RHS and show vice versa, i.e. RHS $\subseteq$ LHS.
  - We show the proof of the first property
    * If $x \in (A \cup B)'$, it means that x does not belong to A or B. In other words x does not belong to A and x does not B either. This means x belongs to the complement of A and to the complement of B, i.e. $x \in A' \cap B'$.
    * Just showing this much does not complete the proof, need to show the other side also.
    * If $x \in A' \cap B'$, it means that x does not belong to A and it does not

belong to B, i.e. it belongs to neither A nor B, i.e. $x \in (A \cup B)'$

  * This completes the argument

- Please read the section on Algebra of Sets, pg 5

Probabilistic models

- There is an underlying process called **experiment** that produces exactly ONE **outcome**.
- A probabilistic model: consists of a sample space and a probability law
  - Sample space (denoted Ω): set of all possible outcomes of an experiment
  - Event: any subset of the sample space
  - Probability Law: assigns a probability to every set A of possible outcomes (event)
  - Choice of sample space (or universe): every element should be distinct and mutually exclusive (disjoint); and the space should be "collectively exhaustive" (every possible outcome of an experiment should be included).

- **Probability Axioms:**
  1. **Nonnegativity.** $P(A) \geq 0$ for every event A.
  2. **Additivity.** If A and B are two **disjoint** events, then
     $P(A \cup B) = P(A) + P(B)$
     (also extends to any countable number of disjoint events).
  3. **Normalization.** Probability of the entire sample space, $P(\Omega) = 1$.
- Probability of the empty set, $P(\phi) = 0$ (follows from Axioms 2 & 3).
- Sequential models, e.g. three coin tosses or two sequential rolls of a die. Tree-based description: see Fig. 1.3
- Discrete probability law: sample space consists of a finite number of possible outcomes, law specified by probability of single element events.
  – Example: for a fair coin toss, $\Omega = \{H, T\}$, $P(H) = P(T) = 1/2$
  – Discrete uniform law for any event A:

$$P(A) = \frac{\text{number of elements in A}}{n}$$

- Continuous probability law: e.g. $\Omega = [0, 1]$: probability of any single element event is zero, need to talk of probability of a subinterval, $[a, b]$ of $[0, 1]$.
  See Example 1.4, 1.5 (This is slightly more difficult. We will cover continuous probability and examples later).

- Properties of probability laws
  1. If $A \subseteq B$, then $P(A) \leq P(B)$
  2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  3. $P(A \cup B) \leq P(A) + P(B)$
  4. $P(A \cup B \cup C) = P(A) + P(A' \cap B) + P(A' \cap B' \cap C)$
  5. Note: book uses A^c for A' (complement of set A).
  6. Proofs: Will be covered in next class. Visualize: Venn diagrams.

Conditional Probability

- Given that we know that an event B has occurred, what is the probability that event A occurred? Denoted by $P(A|B)$. Example: Roll of a 6-sided die. Given that the outcome is even, what is the probability of a 6? Answer: 1/3

- When number of outcomes is finite and all are equally likely,

$$P(A|B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B} \tag{3}$$

- In general,

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)} \tag{4}$$

- $P(A|B)$ is a probability law (satisfies axioms) on the universe B. Exercise: show this.

- Examples/applications
  - Example 1.7, 1.8, 1.11
  - Construct sequential models: $P(A \cap B) = P(B)P(A|B)$. Example: Radar detection (Example 1.9). What is the probability of the aircraft not present and radar registers it (false alarm)?
  - See Fig. 1.9: Tree based sequential description

Total Probability and Bayes Rule

- Total Probability Theorem: Let $A_1, \ldots A_n$ be disjoint events which form a partition of the sample space ($\cup_{i=1}^{n} A_i = \Omega$). Then for any event B,

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \ldots P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + \ldots P(A_n)P(B|A_n) \end{aligned} \tag{5}$$

  Visualization and proof: see Fig. 1.13

- Example 1.13, 1.15

- Bayes rule: Let $A_1, \ldots A_n$ be disjoint events which form a partition of the sample space. Then for any event B, s.t. $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \ldots P(A_n)P(B|A_n)} \tag{6}$$

- Inference using Bayes rule
  - There are multiple "causes" $A_1, A_2, ..A_n$ that result in a certain "effect" B. Given that we observe the effect B, what is the probability that the cause was A_i? Answer: use Bayes rule. See Fig. 1.14
  - Radar detection: what is the probability of the aircraft being present given that the radar registers it? Example 1.16
  - False positive puzzle, Example 1.18: very interesting!

Independence

- $P(A|B) = P(A)$ and so $P(A \cap B) = P(B)P(A)$: the fact that B has occurred gives no information about the probability of occurrence of A. Example: A= head in first coin toss, B = head in second coin toss.
- **“Independence”: DIFFERENT from “mutually exclusive” (disjoint)**
  - Events A and B are disjoint if $P(A \cap B) = 0$: cannot be independent if $P(A) > 0$ and $P(B) > 0$.
    Example: A = head in a coin toss, B = tail in a coin toss
  - Independence: a concept for events in a sequence. Independent events with $P(A) > 0, P(B) > 0$ cannot be disjoint
- Conditional independence **
- Independence of a collection of events

- $P(\cap_{i \in S} A_i) = \Pi_{i \in S} P(A_i)$ for every subset S of $\{1, 2, ..n\}$

- Reliability analysis of complex systems: independence assumption often simplifies calculations
  - Analyze Fig. 1.15: what is $P(\text{system fails})$ of the system $A \rightarrow B$?
    * Let p_i = probability of success of component i.
    * m components in series: $P(\text{system fails}) = 1 - p_1 p_2 \ldots p_m$ (succeeds if all components succeed).
    * m components in parallel: $P(\text{system fails}) = (1 - p_1) \ldots (1 - p_m)$ (fails if all the components fail).

- Independent Bernoulli trials and Binomial probabilities
  - A Bernoulli trial: a coin toss (or any experiment with two possible outcomes, e.g. it rains or does not rain, bit values)
  - Independent Bernoulli trials: sequence of independent coin tosses

– Binomial: Given n independent coin tosses,what is the probability of k heads (denoted $p(k)$)?

  * probability of any one sequence with k heads is $p^k(1-p)^{n-k}$
  * number of such sequences (from counting arguments): $\left(\begin{array}{c} n \\ k \end{array}\right)$
  * $p(k) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}$, where $\left(\begin{array}{c} n \\ k \end{array}\right) \triangleq \frac{n!}{(n-k)!k!}$

– Application: what is the probability that more than c customers need an internet connection at a given time? We know that at a given time, the probability that any one customer needs connection is p.

  Answer: $\displaystyle\sum_{k=c+1}^{n} p(k)$

Counting

- Needed in many situations. Two examples are:
  1. Sample space has a finite number of equally likely outcomes (discrete uniform), compute probability of any event A.
  2. Or compute the probability of an event A which consists of a finite number of equally likely outcomes each with probability p, e.g. probability of k heads in n coin tosses.
- Counting principle (See Fig. 1.17): Consider a process consisting of r stages. If at stage 1, there are n_1 possibilities, at stage 2, n_2 possibilities and so on, then the total number of possibilities = $n_1 n_2 \ldots n_r$.
  - Example 1.26 (number of possible telephone numbers)
  - Counting principle applies even when second stage depends on the first stage and so on, Ex. 1.28 (no. of words with 4 distinct letters)

- Applications: k-permutations.
  - n distinct objects, how many different ways can we pick k objects and arrange them in a sequence?
    * Use counting principle: choose first object in n possible ways, second one in $n-1$ ways and so on. Total no. of ways: $n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$
    * If $k = n$, then total no. of ways = $n!$
    * Example 1.28, 1.29
- Applications: k-combinations.
  - Choice of k elements out of an n-element set without regard to order.
  - Most common example: There are n people, how many different ways can we form a committee of k people? Here order of choosing the k members is not important. Denote answer by $\binom{n}{k}$
  - Note that selecting a k-permutation is the same as first selecting a

k-combination and then ordering the elements (in $k!$) different ways, i.e. $\frac{n!}{(n-k)!} = \binom{n}{k} k!$

- Thus $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- How will you relate this to the binomial coefficient (number of ways to get k heads out of n tosses)?

  Toss number j = person j, a head in a toss = the person (toss number) is in committee

- Applications: k-partitions. **
  - A combination is a partition of a set into two parts
  - Partition: given an n-element set, consider its partition into r subsets of size $n_1, n_2, \ldots, n_r$ where $n_1 + n_2 + \ldots n_r = n$.
    * Use counting principle and k-combinations result.
    * Form the first subset. Choose n_1 elements out of n: $\binom{n}{n_1}$ ways.
    * Form second subset. Choose n_2 elements out of $n - n_1$ available

elements: $\binom{n - n_1}{n_2}$ and so on.

∗ Total number of ways to form the partition:

$$\binom{n}{n_1}\binom{n - n_1}{n_2}\ldots\binom{(n - n_1 - n_2 \ldots n_{r-1})}{n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$$