ECS315 2012/1 Part I.1 Dr.Prapun

1 Probability and You

Whether you like it or not, probabilities rule your life. If you have ever tried to make a living as a gambler, you are painfully aware of this, but even those of us with more mundane life stories are constantly affected by these little numbers.

Example 1.1. Some examples from daily life where probability calculations are involved are the determination of insurance premiums, the introduction of new medications on the market, opinion polls, weather forecasts, and DNA evidence in courts. Probabilities also rule who you are. Did daddy pass you the X or the Y chromosome? Did you inherit grandma's big nose?

Meanwhile, in everyday life, many of us use probabilities in our language and say things like "I'm 99% certain" or "There is a one-in-a-million chance" or, when something unusual happens, ask the rhetorical question "What are the odds?". [15, p 1]

1.1 Randomness

1.2. Many clever people have thought about and debated what randomness really is, and we could get into a long philosophical discussion that could fill up a whole book. Let's not. The French mathematician Laplace (1749–1827) put it nicely:

> "Probability is composed partly of our ignorance, partly of our knowledge."

Inspired by Laplace, let us agree that you can use probabilities whenever you are faced with uncertainty. [15, p 2]

1.3. Random phenomena arise because of [9]:

(a) our partial ignorance of the generating mechanism

(b) the laws governing the phenomena may be fundamentally random (as in quantum mechanics; see also Ex. 1.7.)

(c) our lazy unwillingness to carry out exact analysis because it is not worth the trouble

Example 1.4. Communication Systems [20]: The essence of communication is randomness.

(a) **Random Source**: The transmitter is connected to a random source, the output of which the receiver cannot predict with certainty.

- If a listener knew in advance exactly what a speaker would say, and with what intonation he would say it, there would be no need to listen!

(b) **Noise**: There is no communication problem unless the transmitted signal is disturbed during propagation or reception in a random way.

(c) Probability theory is used to *evaluate the performance* of communication systems.

Example 1.5. Random numbers are used directly in the transmission and security of data over the airwaves or along the Internet.

(a) A radio transmitter and receiver could switch transmission frequencies from moment to moment, seemingly at random, but nevertheless in synchrony with each other.

(b) The Internet data could be credit-card information for a consumer purchase, or a stock or banking transaction secured by the clever application of random numbers.

Example 1.6. Randomness is an essential ingredient in games of all sorts, computer or otherwise, to make for unexpected action and keen interest.

Example 1.7. On a more profound level, quantum physicists teach us that everything is governed by the laws of probability. They toss around terms like the Schrödinger wave equation and Heisenberg's uncertainty principle, which are much too difficult for most of us to understand, but one thing they do mean is that the fundamental laws of physics can only be stated in terms of probabilities. And the fact that Newton's deterministic laws of physics are still useful can also be attributed to results from the theory of probabilities. [15, p 2]

1.8. Most people have preconceived notions of randomness that often differ substantially from true randomness. Truly random data sets often have unexpected properties that go against intuitive thinking. These properties can be used to test whether data sets have been tampered with when suspicion arises. [18, p 191]

- [10, p 174]: "people have a very poor conception of randomness; they do not recognize it when they see it and they cannot produce it when they try"

Example 1.9. Apple ran into an issue with the random shuffling method it initially employed in its iPod music players: true randomness sometimes produces repetition, but when users heard the same song or songs by the same artist played back-to-back, they believed the shuffling wasn't random. And so the company made the feature "less random to make it feel more random," said Apple founder Steve Jobs. [10, p 175]

Please read this

1.2 Background on some frequently used examples

Probabilists love to play with coins and dice. We like the idea of tossing coins, rolling dice, and drawing cards as experiments that have equally likely outcomes.

1.10. ***Coin flipping*** or ***coin tossing*** is the practice of throwing a coin in the air to observe the outcome.

When a **coin** is tossed, it does not necessarily fall heads or tails; it can roll away or stand on its edge. Nevertheless, we shall agree to regard "**head**" (**H**) and "**tail**" (**T**) as the only possible outcomes of the experiment. [3, p 7]

- Typical experiment includes
  - "Flip a coin N times. Observe the sequence of heads and tails" or "Observe the number of heads."

1.11. Historically, ***dice*** is the plural of ***die***, but in modern standard English dice is used as both the singular and the plural. [Excerpted from Compact Oxford English Dictionary.]

- Usually assume six-sided dice
- Usually observe the number of dots on the side facing upwards.

1.12. A complete set of **cards** is called a pack or **deck**.

(a) The subset of cards held at one time by a player during a game is commonly called a **hand**.

(b) For most games, the cards are assembled into a deck, and their order is randomized by **shuffling**.

(c) A standard deck of 52 cards in use today includes thirteen ranks of each of the four French suits.

- The four suits are called spades (♠), clubs (♣), hearts (♡), and diamonds (♢). The last two are red, the first two black.

(d) There are thirteen face values $(2, 3, \ldots, 10$, jack, queen, king, ace) in each suit.

- Cards of the same face value are called of the same **kind**.
- "court" or face card: a king, queen, or jack of any suit.

1.3 A Glimpse at Probability Theory

1.13. Probabilities are used in situations that involve ***randomness***. A ***probability*** is a number used to describe how likely something is to occur, and ***probability*** (without indefinite article) is the study of probabilities. It is "the art of ***being certain of how uncertain you are***." [15, p 2–4] If an event is certain to happen, it is given a probability of 1. If it is certain not to happen, it has a probability of 0. [5, p 66]

1.14. Probabilities can be expressed as fractions, as decimal numbers, or as percentages. If you toss a coin, the probability to get heads is 1/2, which is the same as 0.5, which is the same as 50%. There are no explicit rules for when to use which notation.

- In daily language, proper fractions are often used and often expressed, for example, as "one in ten" instead of 1/10 ("one tenth"). This is also natural when you deal with equally likely outcomes.
- **Decimal numbers** are more common in technical and scientific reporting when probabilities are calculated from data. Percentages are also common in daily language and often with "chance" replacing "probability."
- Meteorologists, for example, typically say things like "there is a 20% chance of rain." The phrase "the probability of rain is 0.2" means the same thing.
- When we deal with probabilities from a theoretical viewpoint, we always think of them as numbers between 0 and 1, not as percentages.
- See also 3.5.

[15, p 10]

Definition 1.15. Important terms [9]:

(a) An activity or procedure or observation is called a **random experiment** if its outcome cannot be predicted precisely because the conditions under which it is performed cannot be predetermined with sufficient accuracy and completeness.

- The term "experiment" is to be construed loosely. We do not intend a laboratory situation with beakers and test tubes.
- Tossing/flipping a coin, rolling a dice, and drawing a card from a deck are some examples of random experiments.

(b) A random experiment may have several separately identifiable **outcomes**. We define the **sample space Ω** as a collection of all possible (separately identifiable) outcomes/results/measurements of a random experiment. Each outcome (ω) is an element, or sample point, of this space.

$\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$

- Rolling a dice has six possible identifiable outcomes $(1, 2, 3, 4, 5,$ and $6)$. $\Omega = \{1,2,3,4,5,6\}$

(c) **Events** are sets (or classes) of outcomes meeting some specifications.

A

- Any[1] event is a subset of Ω.
- Intuitively, an event is a statement about the outcome(s) of an experiment.

$P(A)$

The goal of probability theory is to compute the probability of various events of interest. Hence, we are talking about a *set function* which is defined on subsets of Ω.

Example 1.16. The statement "when a coin is tossed, the probability to get heads is 1/2 (50%)" is a *precise* statement.

(a) It tells you that you are as likely to get heads as you are to get tails.

(b) Another way to think about probabilities is in terms of **average long-term behavior**. In this case, if you toss the coin repeatedly, in the long run you will get *roughly* 50% heads and 50% tails.

[1] For our class, it may be less confusing to allow event A to be any collection of outcomes (, i.e. any subset of Ω).

In more advanced courses, when we deal with uncountable Ω, we limit our interest to only some subsets of Ω. Technically, the collection of these subsets must form a σ-algebra.

Although the outcome of a random experiment is unpredictable, there is a **statistical regularity** about the outcomes. What you cannot be certain of is how the next toss will come up. [15, p 4]

1.17. ***Long-run frequency interpretation***: If the probability of an event A in some actual physical experiment is p, then we believe that if the experiment is *repeated independently* over and over again, then a theorem called the ***law of large numbers*** (LLN) states that, in the long run, the event A will happen approximately $100p\%$ of the time. In other words, if we repeat an experiment a large number of times then the fraction of times the event A occurs will be close to $P(A)$. LLN

Example 1.18. Return to the coin tossing experiment in Ex. 1.16:

Toss a coin	result
1^{st} time	H
2^{nd} time	H
3^{rd} time	T
4^{th} time	H
$\vdots$	

Definition 1.19. Let A be one of the events of a random experiment. If we conduct a sequence of n independent trials of this experiment, and if the event A occurs in $N(A, n)$ out of these n trials, then the fraction

$N(A,1) = 1$
$N(A,2) = 2$
$N(A,3) = 2$
$N(A,4) = 3$

$$\frac{N(A,n)}{n}$$

$\frac{N(A,1)}{1} = 1$
$\frac{N(A,2)}{2} = \frac{2}{2} = 1$
$\frac{N(A,3)}{3} = \frac{2}{3} \approx 0.67$
$\frac{N(A,4)}{4} = \frac{3}{4} = 0.75$

is called the **relative frequency** of the event A in these n trials.

1.20. The long-run frequency interpretation mentioned in 1.17 can be restated as

$$P(A) \text{ ``=''} \lim_{n\to\infty} \frac{N(A,n)}{n}.$$

1.21. Another interpretation: The probability of an outcome can be interpreted as our subjective probability, or degree of belief, that the outcome will occur. Different individuals will no doubt assign different probabilities to the same outcomes.

1.22. In terms of practical range, probability theory is comparable with ***geometry***; both are branches of applied mathematics that are directly linked with the problems of daily life. But while pretty much anyone can call up a natural feel for geometry to some extent, many people clearly have trouble with the development of a good intuition for probability.

- Probability and intuition do not always agree. ***In no other branch of mathematics is it so easy to make mistakes as in probability theory.***
- Students facing difficulties in grasping the concepts of probability theory might find comfort in the idea that even the genius Leibniz, the inventor of differential and integral calculus along with Newton, had difficulties in calculating the probability of throwing 11 with one throw of two dice. (See Ex. 3.4.)

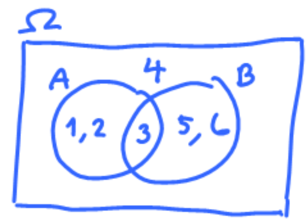
[18, p 4]

2 Review of Set Theory

Example 2.1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 2, 3\}$

$B = \{3, 5, 6\}$

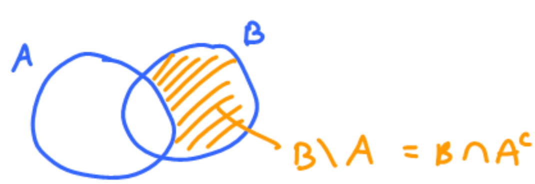


2.2. Venn diagram is very useful in set theory. It is often used to portray relationships between sets. Many identities can be read out simply by examining Venn diagrams.

2.3. If ω is a member of a set A, we write $\omega \in A$. $2 \in A$

Definition 2.4. Basic set operations (set algebra)

- Complementation: $A^c = \{\omega : \omega \notin A\}$. $= \{4, 5, 6\}$
- Union: $A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$ $= \{1, 2, 3, 5, 6\}$
  - Here "or" is inclusive; i.e., if $\omega \in A$, we permit ω to belong either to A or to B or to both.
- Intersection: $A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$ $= \{3\}$
  - Hence, $\omega \in A$ if and only if ω belongs to both A and B.
  - $A \cap B$ is sometimes written simply as AB.
- The ***set difference*** operation is defined by $B \setminus A = B \cap A^c$.
  - $B \setminus A$ is the set of $\omega \in B$ that do not belong to A.
  - When $A \subset B$, $B \setminus A$ is called the complement of A in B.



Reading assignment

2.5. Basic Set Identities:

- Idempotence: $(A^c)^c = A$
- Commutativity (symmetry):

$$A \cup B = B \cup A\ ,\ A \cap B = B \cap A$$

- Associativity:
  - $A \cap (B \cap C) = (A \cap B) \cap C$
  - $A \cup (B \cup C) = (A \cup B) \cup C$
- Distributivity
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- **de Morgan laws**
  - $(A \cup B)^c = A^c \cap B^c$
  - $(A \cap B)^c = A^c \cup B^c$

2.6. Disjoint Sets:

- Sets A and B are said to be ***disjoint*** ($A \perp B$) if and only if $A \cap B = \emptyset$. (They do not share member(s).)

  empty set ; null set

- A collection of sets $(A_i : i \in I)$ is said to be ***pairwise disjoint*** or mutually exclusive [7, p. 9] if and only if $A_i \cap A_j = \emptyset$ when $i \neq j$.

Example 2.7. Sets A, B, and C are pairwise disjoint if

$A \cap B = \phi$,
$A \cap C = \phi$,
and $B \cap C = \phi$

2.8. For a set of sets, to avoid the repeated use of the word "set", we will call it a **collection/class/family** of sets.

Definition 2.9. Given a set S, a collection $\Pi = (A_\alpha : \alpha \in I)$ of subsets[2] of S is said to be a ***partition*** of S if

(a) $S = \bigcup A_{\alpha \in I}$ and

(b) For all $i \neq j$, $A_i \perp A_j$ (pairwise disjoint).

Remarks:

- The subsets A_α, $\alpha \in I$ are called the **parts** of the partition.
- A part of a partition may be empty, but usually there is no advantage in considering partitions with one or more empty parts.

Example 2.10 (Slide:maps)**.**

Example 2.11. Let E be the set of students taking ECS315

A = sets of students who will get A

B+ = " " " " B+

B B

C+ C+

C C

D+ D+

These could be empty { D D

F F

W withdraw }

The collection $\{A, B+, B, C+, C, D+, D, F, W\}$ is a partition of E

Definition 2.12. The ***cardinality*** (or size) of a collection or set A, denoted $|A|$, is the number of elements of the collection. This number may be finite or infinite.

- A *finite* set is a set that has a finite number of elements.
- A set that is not finite is called *infinite*.
- Countable sets:

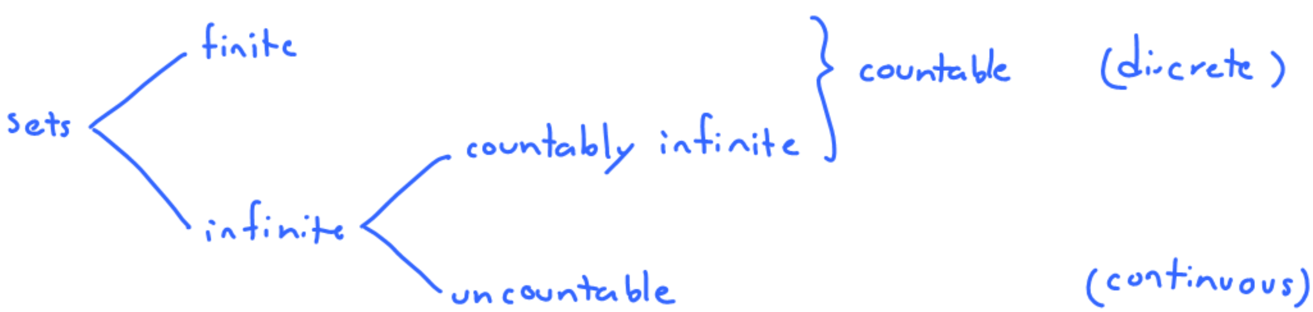
[2]In this case, the subsets are indexed or labeled by α taking values in an index or label set I

- Empty set and finite sets are automatically countable.
- An infinite set A is said to be ***countable*** if the elements of A can be enumerated or listed in a sequence: $a_1, a_2, \ldots$.

infinite + countable = countably infinite

- A ***singleton*** is a set with exactly one element.
  - Ex. $\{1.5\}$, $\{.8\}$, $\{\pi\}$.
  - *Caution*: Be sure you understand the difference between the outcome -8 and the event $\{-8\}$, which is the set consisting of the single outcome -8.

2.13. We can categorize sets according to their cardinality:



Example 2.14. Examples of countably infinite sets:

- the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers,
- the set $\{2k : k \in \mathbb{N}\}$ of all even numbers,
- the set $\{2k - 1 : k \in \mathbb{N}\}$ of all odd numbers,
- the set $\mathbb{Z}$ of integers, $= \{0, +1, -1, +2, -2, \ldots\}$

$\{0, +2, -2, +4, -4, +6, -6, \ldots\}$

$\mathbb{Q}$

Set Theory	Probability Theory
Set	Event
Universal set	Sample Space (Ω)
Element	Outcome (ω)

Table 1: The terminology of set theory and probability theory

	Event Language
A	A occurs
A^c	A does not occur
$A \cup B$	Either A or B occur
$A \cap B$	Both A and B occur

Table 2: Event Language

Example 2.15. Example of uncountable sets[3]:

- $\mathbb{R} = (-\infty, \infty)$
- interval $[0, 1]$ 

- interval $(0, 1]$ 

- $(2, 3) \cup [5, 7)$

Definition 2.16. Probability theory renames some of the terminology in set theory. See Table 1 and Table 2.

- Sometimes, ω's are called states, and Ω is called the state space.

2.17. Because of the mathematics required to determine probabilities, probabilistic methods are divided into two distinct types, discrete and continuous. A discrete approach is used when the number of experimental outcomes is finite (or infinite but countable). A continuous approach is used when the outcomes are continuous (and therefore infinite). It will be important to keep in mind which case is under consideration since otherwise, certain paradoxes may result.

[3]We use a technique called diagonal argument to prove that a set is not countable and hence uncountable.

3 Classical Probability

Classical probability, which is based upon the ratio of the number of outcomes favorable to the occurrence of the event of interest to the total number of possible outcomes, provided most of the probability models used prior to the 20th century. It is the first type of probability problems studied by mathematicians, most notably, Frenchmen Fermat and Pascal whose 17th century correspondence with each other is usually considered to have started the systematic study of probabilities. [15, p 3] Classical probability remains of importance today and provides the most accessible introduction to the more general theory of probability.

Definition 3.1. Given a finite sample space Ω, the ***classical probability*** of an event A is

$$P(A) = \frac{|A|}{|\Omega|} \tag{1}$$

[4, Defn. 2.2.1 p 58]. In traditional language, a probability is a fraction in which the bottom represents the number of possible outcomes, while the number on top represents the number of outcomes in which the event of interest occurs.

- Assumptions: When the following are not true, do not calculate probability using (1).

equipossible
equiprobable
equally likely
fair

  - Finite Ω: The number of possible outcomes is finite.
  - Equipossibility: The outcomes have equal probability of occurrence.

- The bases for identifying equipossibility were often
  - physical symmetry (e.g. a well-balanced dice, made of homogeneous material in a cubical shape) or
  - a balance of information or knowledge concerning the various possible outcomes.

- Equipossibility is meaningful only for finite sample space, and, in this case, the evaluation of probability is accomplished through the definition of classical probability.

and the next section

- We will NOT use this definition beyond this section. We will soon introduce a formal definition in Section 5.

- In many problems, when the finite sample space does not contain equally likely outcomes, we can redefine the sample space to make the outcome equipossible.

10,000 balls 9,999 black
1 red

$\Omega = \cancel{\{red, black\}}$

$\Omega = \{10{,}000 \text{ balls}\}$

Grab one ball.

Example 3.2 (Slide)**.** In drawing a card from a deck, there are 52 equally likely outcomes, 13 of which are diamonds. This leads to a probability of 13/52 or 1/4.

3.3. Basic properties of classical probability: From Definition 3.1, we can easily verified[4] the properties below.

- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ which comes directly from
$$|A \cup B| = |A| + |B| - |A \cap B|.$$
- $A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$
- Suppose $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $P(\{\omega_i\}) = \frac{1}{n}$. Then $P(A) = \sum\limits_{\omega \in A} P(\{\omega\})$.
    - The probability of an event is equal to the sum of the probabilities of its component outcomes because outcomes are mutually exclusive

[4] Because we will not rely on Definition 3.1 beyond this section, we will not worry about how to prove these properties. In Section 5, we will prove the same properties in a more general setting.

Leibniz

Example 3.4 (Slides)**.** When rolling two dice, there are 36 (equiprobable) possibilities.

$$P[\text{sum of the two dice} = 5] = 4/36.$$

Though one of the finest minds of his age, Leibniz was not immune to blunders: he thought it just as easy to throw 12 with a pair of dice as to throw 11. The truth is...

$$P[\text{sum of the two dice} = 11] = 2/36$$
$$P[\text{sum of the two dice} = 12] = 1/36$$

Definition 3.5. In the world of gambling, probabilities are often expressed by ***odds***. To say that the odds are n:1 *against* the event A means that it is n times as likely that A does not occur than that it occurs. In other words, $P(A^c) = nP(A)$ which implies $P(A) = \frac{1}{n+1}$ and $P(A^c) = \frac{n}{n+1}$.

"Odds" here has nothing to do with even and odd numbers. The odds also mean what you will win, in addition to getting your stake back, should your guess prove to be right. If I bet \$1 on a horse at odds of 7:1, I get back \$7 in winnings plus my \$1 stake. The bookmaker will break even in the long run if the probability of that horse winning is 1/8 (not 1/7). Odds are "even" when they are 1:1 - win \$1 and get back your original \$1. The corresponding probability is 1/2.

3.6. It is important to remember that classical probability relies on the assumption that the outcomes are ***equally likely***.

Example 3.7. ***Mistake*** made by the famous French mathematician Jean Le Rond d'Alembert (18th century) who is an author of several works on probability:

"The number of heads that turns up in those two tosses can be 0, 1, or 2. Since there are three outcomes, the chances of each must be 1 in 3." ← incorrect

Correction : Four possible outcomes

HH

HT

TH

TT

$P[\#H = 0] = \frac{1}{4} = P[\#H = 2]$

$P[\#H = 1] = \frac{2}{4} = \frac{1}{2}$