ECE 361 Probability for Engineers (Fall, 2016)
Lecture 1a

§1.1 Sets

A set S is an **unordered** collection of **unique** objects, called elements, with membership of element x in S denoted $x \in S$.

- *Empty* (or null) set: $\emptyset$.
- Not a member of S: $x \notin S$.
- Finite set of n elements: $S = \{x_1, \ldots, x_n\}$.
- *Countably infinite* set $S = \{x_1, x_2, \ldots, \}$, also written $S = \{x_n, n \in \mathbb{N}\}$, for $\mathbb{N} = \{1, 2, 3, \ldots\}$ the natural numbers.
- Uncountably infinite set $S = [0, 1]$, where the notation $[a, b]$ denotes the interval of all real numbers between (and including) a and b. The open interval (a, b) denotes all points between a and b, but excluding a and b. S *uncountably infinite* means S cannot be put into a one to one correspondence with the natural numbers.
- Those elements that satisfy property P: $\{x : x \text{ satisfies } P\}$.
- *Equal sets* have identical elements: $S = T$ means $x \in S \Leftrightarrow x \in T$.
- *Universe* Ω: the largest relevant set of interest.
- *Subsets*: $A \subseteq B$ means if $x \in A$ then $x \in B$, i.e., each element in A is also an element of B. $A \subset B$ (a strict subset) means A is a subset of B and moreover at least one element of B is *not* in A
- *Superset*: $A \supset B$ means $B \subset A$.
- *Set difference*: $A \setminus B$ (sometimes written $A - B$) are elements in A not in B: $A \setminus B = \{x \in A \text{ and } x \notin B\}$. The *symmetric difference*, often denoted $A \Delta B$ is $(A \setminus B) \cup (B \setminus A)$, i.e., those elements in A or B but not in both.

The set $\{1, \ldots, n\}$, i.e., the set of integers 1 through n, is denoted $[n]$.

Set operations

- *Set complement* (with respect to universe Ω): $S^c = \{x : x \notin S\}$. The complement of the universe is the empty set $\Omega^c = \emptyset$, i.e., the set of objects not in the universe is the null set.
- *Set union*: $S \cup T = \{x : x \in S \text{ or } x \in T\}$. Union of a finite collection of subsets $(S_1, \ldots, S_N)$ of Ω:
$$\bigcup_{n=1}^{N} S_n = S_1 \cup S_2 \cup \cdots \cup S_n = \{x : x \in S_n \text{ for some } n \in [N]\}. \tag{1}$$
- *Set intersection*: $S \cap T = \{x : x \in S \text{ and } x \in T\}$. Intersection of a finite collection of subsets $(S_1, \ldots, S_N)$ of Ω:
$$\bigcap_{n=1}^{N} S_n = S_1 \cap S_2 \cap \cdots = \{x : x \in S_n \text{ for all } n \in [N]\}. \tag{2}$$
- *Disjoint sets* have no elements in common: $S \cap T = \emptyset$. A collection of subsets $(S_1, \ldots, S_N)$ of Ω are pairwise disjoint if $S_i \cap S_j = \emptyset$ for each $i, j \in [N]$, i.e., each pair of subsets has no elements in common.
- *Set cover.* A collection of subsets $(S_1, \ldots, S_N)$ of Ω covers Ω if $\bigcup_{n=1}^{N} S_n = \Omega$.
- *Set partition.* A partition of sets is a collection of pairwise disjoint subsets that cover Ω:
$$\{S_1, \ldots, S_n\} \text{ partitions } \Omega \Leftrightarrow \bigcup_{i=1}^{n} S_i = \Omega \text{ and } S_i \cap S_j = \emptyset, \forall i, j \in [n]. \tag{3}$$
- *Cardinality.* Given a finite set S, let $|S|$ denote its cardinality, i.e., the number of elements in the set. Example: $|\{x_1, \ldots, x_n\}| = n$. Example: $|[n]| = n$.
- *Power set.* Given a finite set S, its power set, denoted $\mathcal{P}(S)$ or 2^S, is the set of all subsets of S. For example, $S = \{H, T\}$ then $\mathcal{P}(S) = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$. Note if $|S| = n$ then $|2^S| = 2^n$. Easy proof: each subset of S can be constructed by considering each element $x \in S$ in turn, and either including it or not. There are 2^n distinct subsets under this operation. Note elements of $\mathcal{P}(S)$ are subsets of S. Thus: $T \subset S \Leftrightarrow T \in \mathcal{P}(S)$.

The algebra of sets

Venn diagrams are useful for understanding the following basic consequences of the above set operations and definitions.

$$S \cup T = T \cup S \qquad S \cup (T \cup U) = (S \cup T) \cup U$$
$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U) \qquad S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$$
$$(S^c)^c = S \qquad S \cap S^c = \emptyset \tag{4}$$
$$S \cup \Omega = \Omega \qquad S \cap \Omega = S \tag{5}$$

De Morgan's laws. Given a collection of subsets $(S_1, \ldots, S_N)$ of Ω:

$$\left(\bigcup_{n \in [N]} S_n \right)^c = \bigcap_{n \in [N]} S_n^c, \quad \left(\bigcap_{n \in [N]} S_n \right)^c = \bigcup_{n \in [N]} S_n^c. \tag{6}$$

The first says the complement of the union of the collection is the intersection of the complements. Both describe the set of elements not found in any of the subsets S_n.

§1.2 Probabilistic models

Useful definition from text: "A probabilistic model is a mathematical description of an uncertain situation".

Think of these uncertain situations as random experiments. They are random in that the outcome is uncertain and is likely to vary upon independent trials of the experiment.

The probability triple $(\Omega, \mathcal{F}, \mathbb{P})$:

- *Sample space* (Ω). An outcome of the random experiment is denoted ω, and the set of possible outcomes is the universe Ω. The random experiment produces precisely one outcome $\omega \in \Omega$. This set may be finite or countably infinite.
- *Events* ($\mathcal{F}$). An event $A \in \mathcal{F}$ is a subset of outcomes $A \subset \Omega$. Therefore the set of all possible events is the power set of outcomes: $\mathcal{F} = \mathcal{P}(\Omega)$.
- *Probability* ($\mathbb{P}$): a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ assigning a number $\mathbb{P}(A) \in [0, 1]$ to each $A \in \mathcal{F}$, satisfying the axioms of probability given later in this section. $\mathbb{P}(A)$ is the probability of event A being true, i.e., the probability that the random outcome of the experiment, ω, lies in A: $\mathbb{P}(\omega \in A)$.

The restriction to finite or countably infinite sample space Ω is sufficient for our purposes until we get to continuous random variables later in the course. The probability triple for continuous sample spaces is slightly more involved.

Sample spaces and events

A random experiment produces an outcome $\omega \in \Omega$. Events are subsets of outcomes, $A \subset \Omega$. The set of possible events $\mathcal{F}$ is the power set of Ω, $\mathcal{F} = \mathcal{P}(\Omega)$.

A few canonical examples:

- Toss 1 coin: $\Omega = \{H, T\}$
- Toss of 2 ordered coins: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$.
- Toss of 2 unordered coins: $\Omega = \{\{H, H\}, \{T, H\}, \{T, T\}\}$.
- Toss a coin until a head comes up and record the number of required tosses: $\Omega = \{1, 2, 3, \ldots, \} = \mathbb{N}$.
- The color of a traffic light $\Omega = \{\text{red}, \text{yellow}, \text{green}\}$.
- The genders of two children $\Omega = \{\text{both male}, \text{both female}, \text{one male and one female}\}$.
- The grades assigned in a class $\Omega = \{A+, A, A-, B+, B, B-, C+, C, C-, D+, D, F\}$.

Choosing an appropriate sample space

The sample space should be as simple as possible but no simpler than necessary. Choose a sample space that is:

- *Collectively exhaustive:* the sample space should be rich enough to ensure there always **exists** (at least one) an outcome corresponding to the result of the experiment. *Example:* if a coin may land on its side then $\Omega = \{\text{head}, \text{tail}, \text{side}\}$.
- *Mutually exclusive:* the random experiment should produce a **unique** (at most one) outcome in the sample space. *Example:* suppose the coin may roll off the table on to the floor, and we set $\Omega = \{\text{head}, \text{tail}, \text{side}, \text{floor}\}$. But this set is not mutually exclusive since a coin on the floor will show heads, tails, or lie on its side.

Sequential models

Many random experiments have a sequential nature to them. Tree-based models are often useful for describing the outcomes of such experiments. See Fig. 1.3 in text. Some canonical examples:

- Toss of 2 ordered coins: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$.
- Toss a coin then roll a die: $\Omega = \{(C,D) : C \in \{H,T\}, D \in [6]\}$.

Observe the important notational convention that $\{A,B,C\}$ is the *unordered* set with distinct elements A,B,C, while (A,B,C) is the *ordered* set with (possibly nondistinct) elements A,B,C. Thus $\{A,B,A\}$ is not allowed, but (A,B,A) is perfectly valid.

Probability laws

The probability $\mathbb{P} : \mathcal{F} \to [0,1]$ in the probability triple $(\Omega,\mathcal{F},\mathbb{P})$ modeling a random experiment must obey the following three axioms of probability.

Probability axioms:

- Nonnegativity: $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
- Additivity: if A,B are disjoint events ($A \cap B = \emptyset$) then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. More generally if $|\Omega| = \infty$ and $\{A_1, A_2, \ldots\}$ is an infinite sequence of disjoint events, then $\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$.
- Normalization: $\mathbb{P}(\Omega) = 1$ (something must happen).

The *additivity* property tells us how to assign the probability $\mathbb{P}(A)$ for an event $A \subseteq \Omega$, given the probabilities $\mathbb{P}(\{\omega\})$ for each outcome $\omega \in \Omega$:

Discrete probability law. If Ω is finite then the probability axioms are satisfied by assigning a probability to each outcome ω (the atomic or singleton events) such that $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$. The probability of any event $A \subseteq \Omega$ is then given by

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega). \tag{7}$$

Such an assignment will satisfy the probability axioms for all events $A \in \mathcal{F}$.

That is, to compute the probability of the event A we simply *add up* the probabilities of each outcome in the event. That this is valid follows from the *mutually exclusive* property of the set of outcomes.

If all outcomes are equally likely then we obtain the discrete uniform probability law.

Discrete uniform probability law. If $|\Omega| = n$ and $\mathbb{P}(\{\omega\}) = 1/n$ for all $\omega \in \Omega$ (all outcomes are equally likely), then

$$\mathbb{P}(A) = \frac{|A|}{n}, \ \forall A \in \mathcal{F}. \tag{8}$$

§1.2 Probabilistic models

Discrete models

Assigning probabilities to events can be done in an intuitive way for many cases. Here are some examples:

1. Single toss of a fair coin. Set $\Omega = \{H, T\}$. Observe $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$. We must *assign* a probability to each element of $\mathcal{F}$, i.e., each subset of Ω:
   - As the coin is fair, we *assign* $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 1/2$.
   - By additivity: $\mathbb{P}(\Omega) = \mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1/2 + 1/2 = 1$ (as required by *normalization*).
   - As $A \cap \emptyset = \emptyset$ we have $\mathbb{P}(A \cup \emptyset) = \mathbb{P}(A) = \mathbb{P}(A) + \mathbb{P}(\emptyset)$, and thus $\mathbb{P}(\emptyset) = 0$.

   This probability assignment adheres to the three axioms (check!).

2. Single toss of a biased coin. Again set $\Omega = \{H, T\}$ and thus $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$. We must *assign* a probability to each element of $\mathcal{F}$:
   - Suppose we know or believe the probability of a head is p for some number $p \in [0, 1]$; we therefore *assign* $\mathbb{P}(\{H\}) = p$.
   - As $\{H\}, \{T\}$ are disjoint and $\{H\} \cup \{T\} = \Omega$, we have $\mathbb{P}(\{H\} \cup \{T\}) = \mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = \mathbb{P}(\Omega) = 1$, and this equation lets us conclude $\mathbb{P}(\{T\}) = 1 - p$.
   - Critically, not all outcomes are equally likely, so the *discrete uniform probability law* does *not* apply.
   - As before, and as is true in general $\mathbb{P}(\emptyset) = 0$.

   This probability assignment adheres to the three axioms (check!).

3. Toss three three fair coins: a penny, a nickle, and a dime, and record their three faces.
   - Now $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, and $\mathcal{F} = \mathcal{P}(\Omega)$ has $2^8 = 256$ elements.
   - We will *assign* a probability to each element of Ω, and then, by additivity, the probability for any $A \in \mathcal{F}$ will be computed as $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.
   - Each outcome is equally likely with probability 1/8.
   - Example: consider $A = \{\text{exactly two heads occur}\}$. Then, since all outcomes are equally likely, by the discrete uniform probability law:
   $$\mathbb{P}(A) = \mathbb{P}(\{HHT, HTH, THH\}) = \frac{|A|}{|\Omega|} = 3/8. \tag{9}$$

4. Consider the experiment of rolling a pair of *distinct* fair four-sided dice. As the dice are distinct,
   $$\Omega = \{(i, j) : i \in [4], j \in [4]\} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} \;\;1\;\; & \;\;\;2\;\;\; & \;\;\;3\;\;\; & \;\;\;4\;\; \end{array} \\ \begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,1) & (2,2) & (2,3) & (2,4) \\ (3,1) & (3,2) & (3,3) & (3,4) \\ (4,1) & (4,2) & (4,3) & (4,4) \end{bmatrix} \end{array}, \tag{10}$$

   Here the rows correspond to the face of the first die, and the columns correspond to the face of the second die. Fairness means all outcomes are equally likely, so $\mathbb{P}(\omega) = 1/16$ for each $\omega = (i, j)$. The probabilities of various events of possible interest are then given by:
   - $\mathbb{P}(\text{sum is even}) = 8/16$
   $$\{\text{sum is even}\} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \times & & \times & \\ & \times & & \times \\ \times & & \times & \\ & \times & & \times \end{bmatrix} \end{array} \tag{11}$$

- $\mathbb{P}(\text{dice show the same face}) = 1/4$

$$\{\text{dice show the same face}\} = \begin{array}{cc} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \times & & & \\ & \times & & \\ & & \times & \\ & & & \times \end{bmatrix} \end{array} \quad (12)$$

- $\mathbb{P}(\text{first is larger than second}) = 6/16$

$$\{\text{first is larger than second}\} = \begin{array}{cc} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} & & & \\ \times & & & \\ \times & \times & & \\ \times & \times & \times & \end{bmatrix} \end{array} \quad (13)$$

- $\mathbb{P}(\text{at least one roll is a } 4) = 7/16$

$$\{\text{at least one roll is a } 4\} = \begin{array}{cc} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} & & & \times \\ & & & \times \\ & & & \times \\ \times & \times & \times & \times \end{bmatrix} \end{array} \quad (14)$$

5. Repeat with a pair of *indistinguishable* fair four-sided dice. Now

$$\Omega = \begin{bmatrix} \{1,1\} & \{1,2\} & \{1,3\} & \{1,4\} \\ & \{2,2\} & \{2,3\} & \{2,4\} \\ & & \{3,3\} & \{3,4\} \\ & & & \{4,4\} \end{bmatrix} \quad (15)$$

The $|\Omega| = 10$ outcomes are *not* equally probable. The outcomes on the diagonal each have probability $1/16$, while the outcomes above the diagonal each have probability $1/8$: observe this yields $\mathbb{P}(\Omega) = 1$. Probabilities of the above events under this new sample space are:

- $\mathbb{P}(\text{sum is even}) = 8/16$

$$\{\text{sum is even}\} = \begin{bmatrix} \times & & \times & \\ & \times & & \times \\ & & \times & \\ & & & \times \end{bmatrix} \quad (16)$$

- $\mathbb{P}(\text{dice show the same face}) = 1/4$

$$\{\text{dice show the same face}\} = \begin{bmatrix} \times & & & \\ & \times & & \\ & & \times & \\ & & & \times \end{bmatrix} \quad (17)$$

- $\mathbb{P}(\text{first is larger than second})$. This question is ill-posed as the dice are assumed to be *indistinguishable*, so there is no first and second.
- $\mathbb{P}(\text{at least one roll is a } 4) = 7/16$

$$\{\text{at least one roll is a } 4\} = \begin{bmatrix} & & & \times \\ & & & \times \\ & & & \times \\ & & & \times \end{bmatrix} \quad (18)$$

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.