Math/Stats 425
Introduction to Probability

1. Uncertainty and the axioms of probability

- Processes in the real world are **random** if outcomes cannot be predicted with certainty.
- Example: coin tossing, stock market prices, weather prediction, genetics, games of chance, industrial quality control, estimating a movie rating based on previous customer preferences, internet packages arriving at a router, etc.
- Humans have fairly poorly developed intuitions about probabilities, so a **formal mathematical approach** is necessary. It can lead to surprising results in some simple problems.

Example: The birthday problem

What is the chance that 2 or more people in a room share a birthday?

# people	25	40	60	70
probability	0.57	0.891	0.994	0.999

Example: HIV testing

Suppose 99% of people with HIV test positive, 95% of people without HIV test negative, and 0.1% of people have HIV. What is the chance that a person who tests positive actually has HIV?

Answer: $\approx \frac{1}{51} < 2\%$.

Outcomes, Sample Spaces & Events

- A random process (or "experiment") gives rise to an **outcome**.
- The first task in specifying a mathematical model for a random process is to define the **sample space**, $\mathbb{S}$, which is the set of all possible outcomes.

Example 1. A coin can land heads or tails, so we set $\mathbb{S} = \{H, T\}$.

Note: this is a mathematical idealization. A coin can also land on its edge, or roll away and get lost. We simplify the problem, much as in mechanics we might assume a planet is a point mass, or a pulley has no friction.

Example 2. Roll a die 3 times. If we take the outcome to be a sequence of 3 numbers,

$$\mathbb{S} = \{111, 112, 113, \ldots, 666\},$$

the total number of outcomes is

$$\#\mathbb{S} =$$

Example 3. Roll a die 3 times. Taking the outcome to be the sum of the 3 rolls,

$$\begin{aligned} \mathbb{S} &= \{3, 4, 5, \ldots, 18\} \\ \#\mathbb{S} &= \end{aligned}$$

In different contexts, you can choose different sample spaces for the same experiment.

Example 4. Toss a coin until heads appears:

$$\mathbb{S} = \{H, TH, TTH, TTTH, \ldots\}.$$

Here, $\#\mathbb{S} = \infty$, with $\mathbb{S}$ being **countably infinite**.

Example 5. Measure the lifetime of an LED lightbulb:

$$\mathbb{S} = [0, \infty) = \{t : 0 \leq t < \infty\}.$$

Here, $\mathbb{S}$ is **uncountable**.

- An **event**, E, is a subset of $\mathbb{S}$. The event E **occurs** if the process produces an outcome in E.
- Events can be written in different ways. For example, suppose we roll a die 3 times and take as our sample space $\mathbb{S} = \{111, 112, \ldots, 666\}$. We can define E to be the event that the sum is even. Equivalently, we can write

$$E = \{\text{the sum of 3 dice is even}\} \tag{1}$$

or

$$E = \{112, 114, 116, 121, 123, 125, \ldots, 666\} \tag{2}$$

- Note that (1) and (2) are only equivalent for this choice of $\mathbb{S}$.
- Now take $F = \{566, 656, 665, 666\}$. How could you write F in words?

Using set operations on events

- The **complement** E^c, sometimes written $\bar{E}$, is the set of outcomes that are not in E. We say E^c is the event that E does not happen (happen means the same thing as occur).

- The **intersection** $E \cap F$, sometimes written EF, is the event that E **and** F occur. Formally, $E \cap F$ is the set of outcomes in both E and F.

- The **union** $E \cup F$ is the event that E **or** F occurs. Formally, $E \cup F$ is the set of outcomes in either E or F or both.

Discussion Problem. A box contains 3 marbles, one red, one green, one blue. A marble is drawn, then replaced, and then a second is drawn.

(i) Describe the sample space (both in words and by listing outcomes). How large is it?

(ii) Let $E = \{\text{a red marble is drawn at least once}\}$ and $F = \{\text{a blue marble is drawn at least once}\}$. Describe the events $E \cap F$ and $E \cup F$, and list their outcomes.

More set notation

- E is a **subset** of F, written $E \subset F$ or $F \supset E$, if all outcomes in E are also in F. If $E \subset F$, occurrence of E implies the occurrence of F.

- The **empty set** $\emptyset$ is the set with no elements, also known as the **null event**.

  Example 1. If $\mathbb{S} = \{H, T\}$, $E = \{H\}$, $F = \{T\}$, then $E \cap F = \emptyset$.

  Example 2. $E \cap E^c = \emptyset$ for any event E.

- Events E and F are **mutually exclusive**, or **disjoint**, if $E \cap F = \emptyset$.

  Example 3. The empty set is disjoint to itself, since $\emptyset \cap \emptyset = \emptyset$.

- E and F are **identical** if E occurs whenever F occurs, and vice versa. We then write $E = F$. Formally, $E = F$ means $E \subset F$ and $F \subset E$.

Venn diagrams

- These are a good way to visualize relationships between events, and to find simple identities.

Example. Using a Venn diagram, write $E \cup F$ as a union of mutually exclusive events.

Formal properties for set operations

- **Commutative**:

$E \cup F = F \cup E$,

$E \cap F = F \cap E$.

- **Associative**:

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

$(E \cap F) \cap G = E \cap (F \cap G)$.

- **Distributive**:

$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$,

$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

- **DeMorgan's law**:

$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$,

$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$.

“Long run” interpretation of probability

- Many random processes or experiments are repeatable. We can then assign probabilities empirically.

- Perform an experiment n times. Let $n(E)$ denote the number of times that E occurs. Then the probability of E should satisfy

  $\mathbb{P}(E) \approx \dfrac{n(E)}{n}$, the proportion of time E occurs.

  Example: coin tossing. We could say $\mathbb{P}(H) = 1/2$ since, in a large number of tosses, the proportion of heads is $\approx 1/2$.

- For non-repeatable events (e.g. the Dow Jones index on 1/10/2013) we can imagine repetitions (“parallel universes”).

- The long run view of probability has limited practical use, as n is often not large. As J. M. Keynes observed, “In the long run, we are all dead.”

Properties of "long run" probabilities

- Since $0 \leq n(E) \leq n$, we have $0 \leq \dfrac{n(E)}{n} \leq 1$.

- $n(\mathbb{S}) = n$ and so $\dfrac{n(\mathbb{S})}{n} = 1$.

- If E and F are disjoint events then $n(E \cup F) = n(E) + n(F)$ and so
$$\frac{n(E \cup F)}{n} = \frac{n(E)}{n} + \frac{n(F)}{n}.$$

- **Remarkably, if we require that probabilities must satisfy these three "long run" properties, we get a mathematical foundation for the entire theory of probability.**

The axioms of probability

Consider an experiment with sample space $\mathbb{S}$. Write $\mathbb{P}(E)$ for the probability of an event E. We assume that $\mathbb{P}$ obeys the following axioms:

A1: For any event E, $\boxed{0 \leq \mathbb{P}(E) \leq 1}$

A2: $\boxed{\mathbb{P}(\mathbb{S}) = 1}$

A3: If $E_1, E_2, \ldots$ is a countably infinite sequence of **mutually exclusive** events (i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j$) then $\boxed{\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)}$

- All mathematical understanding of probability is built on these three axioms.

> "The theory of probability is the only mathematical tool available to help map the unknown and the uncontrollable. It is fortunate that this tool, while tricky, is extraordinarily powerful and convenient."
>
> **Benoit Mandelbrot**, *The Fractal Geometry of Nature (1977)*

Some Consequences of the Axioms

Proposition 1. $\boxed{\mathbb{P}(\emptyset) = 0}$

Proof:

Proposition 2. (A finite version of axiom A3)

If $E_1, \ldots, E_n$ are mutually exclusive then

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

Proof:

Proposition 3. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$

Proof:

Proposition 4. If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.

Proof: by decomposing F as a disjoint union.

Proposition 5. (Addition rule)

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

Note: This is the general addition rule. If E and F are mutually exclusive, then $\mathbb{P}(E \cap F) = 0$ and we recover Prop. 2.

Proof: by decomposition into a disjoint union.

Discussion Problem. A randomly selected student has a VISA card with probability 0.61, a Mastercard with probability 0.24, and both with probability 0.11. Find the chance that a random student has neither VISA nor Mastercard.

Proposition 6. (Addition rule for 3 events).

$$\begin{aligned}\mathbb{P}(E \cup F \cup G) = {} & \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) \\ & -\mathbb{P}(E \cap F) - \mathbb{P}(F \cap G) - \mathbb{P}(G \cap E) \\ & +\mathbb{P}(E \cap F \cap G)\end{aligned}$$

Sketch proof:

How does this extend to $\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n)$? (Not required for this course; see Ross, p. 30, Prop. 4.4.)

The classical probability model

- By symmetry or other considerations, it may be reasonable to suppose that each outcome in $\mathbb{S}$ is equally likely.
- Suppose $\mathbb{S} = \{s_1, \ldots, s_N\}$. If outcomes are equally likely then axiom A1 implies that $\mathbb{P}(\{s_i\}) = \frac{1}{N}$ for $i = 1, \ldots, N$.
- It follows that for any event E

$$\boxed{\mathbb{P}(E) = \frac{\text{number of outcomes in } E}{\text{total number of outcomes in } \mathbb{S}}}$$

Why?

- Classical probability problems then become problems of counting.

Example. I roll two dice. What is the chance that the sum will be 6?

Solution. Suppose that one die is red and the other is black. Now take $\mathbb{S}$ to list pairs of outcomes for the red and black die respectively, so $\mathbb{S} = \{11, 12, 13, \ldots, 66\}$.

Note: the symmetry argument for the classical probability model fails if we set $\mathbb{S} = \{2, 3, \ldots, 12\}$.