Lecture 3: Aug 25

Last time

- Set theory (1.1)
- Axiomatic Foundations (1.2)

Today

- Axiomatic Foundations (1.2)
- Calculus of Probabilities (1.2)
- Conditional Probability (1.3)

Theorem For any three events, A, B, and C, defined on a sample space S,

1. Commutativity
$$A \cup B = B \cup A,$$
$$A \cap B = B \cap A;$$

2. Associativity
$$A \cup (B \cup C) = (A \cup B) \cup C,$$
$$A \cap (B \cap C) = (A \cap B) \cap C;$$

3. Distributive Laws
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$$

4. DeMorgan's Laws
$$(A \cup B)^c = A^c \cap B^c,$$
$$(A \cap B)^c = A^c \cup B^c;$$

We show the proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ in the distributive laws. Caution: Venn diagrams are helpful in visualization, but they do not constitute a formal proof. To prove that two sets are equal, we need to show that each set contains the other.
proof:

- $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$:
  Let $x \in (A \cap (B \cup C))$. By definition of intersection, $x \in (B \cup C)$ that is, either $x \in B$ or $x \in C$. Since x also must be in A, we have that either $x \in (A \cap B)$ or $x \in (A \cap C)$; therefore, $x \in ((A \cap B) \cup (A \cap C))$.
- $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$:
  Let $x \in ((A \cap B) \cup (A \cap C))$. This implies that $x \in (A \cap B)$ or $x \in (A \cap C)$. If $x \in (A \cap B)$, then x is in both A and B. Since $x \in B$, then $x \in (B \cup C)$and thus

$x \in (A \cap (B \cup C))$. It follows the same argument when $x \in (A \cap C)$, we still have $x \in (A \cap (B \cup C))$.

Definition Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \varnothing$. The events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Definition If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots = S$, then the collection of $A_1, A_2, \ldots$ forms a *partition* of S.

Example The sets $A_i = [i, i+1), i = 0, 1, 2, \ldots$ form a partition of $[0, \infty)$.

Basics of Probability Theory

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, then

- different outcomes may occur each time
- some outcomes may repeat
- the "frequency of occurrence" of an outcome can be thought of as a probability

However, we **do not** define probabilities in terms of frequencies but instead take the mathematically simpler axiomatic approach. The axiomatic approach is not concerned with the interpretations of probabilities, but is concerned only that the probabilities are defined by a function satisfying the axioms. Interpretations of the probabilities are quite another matter:

- The "frequency of occurrence" of an event is one example of a particular interpretation of probability.
- Another possible interpretation is a subjective one, where we can think of the probability as a belief in the chance of an event occurring.

Axiomatic Foundations

For each event A in the sample space S, we want to associate with A a number between zero and one that will be called the probability of A, denoted by $\Pr(A)$. The domain of Pr is the set where the arguments of the function $\Pr(\dot{})$ are defined. It is natural to define the domain of Pr as all subsets of S, that is for each $A \subset S$, we define $\Pr(A)$ as the probability that A occurs. However, there are some technical difficulties to overcome which requires us to familiarize with the following.

Definition A collection of subsets of S is called a *sigma algebra* (or *Borel field*), denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\varnothing \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$).
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation).

3. If $A_1, A_2, \cdots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

From Property (1) and (2), we see that the empty set and its complement S (since $S = \varnothing^c$) are always in a sigma algebra. In fact, they construct the *trivial* algebra $\{\varnothing, S\}$ which is the smallest sigma algebra.

By DeMorgan's Law, (3) can be replaced by:

$$3'.\ \text{if } A_1, A_2, \cdots \in \mathcal{B}, \text{ then } \cap_{i=1}^{\infty} A_i \in \mathcal{B}.$$

This is because:

Example If S is finite or countable (where the elements of S can be put into $1-1$ correspondence with a subset of the integers), then these technicalities really do not arise, for we we define for a given sample space S,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathcal{B}$ (why?).[hint: for each element, it is either in or out of a subset, so 2 choices].

Example Let $S = (-\infty, \infty)$, the real line. Then $\mathcal{B}$ is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathcal{B}$, it follows that $\mathcal{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

Definition Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function Pr with domain $\mathcal{B}$ that satisfies

1. $\Pr(A) \geqslant 0$ for all $A \in \mathcal{B}$.
2. $\Pr(S) = 1$.
3. If $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

Example Consider the simple experiment of tossing a fair coin (just once), so $S = \{H, T\}$. A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since $S = \{H\} \cup \{T\}$, we have , from Axiom 1, $\Pr(\{H\} \cup \{T\}) = 1$. Also, $\{H\}$ and $\{T\}$ are disjoint, so $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$. Collectively, we have

$$\begin{aligned}
&\Pr(\{H\}) = \Pr(\{T\}) \\
&\Pr(\{H\} \cup \{T\}) = 1 \\
&\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})
\end{aligned}$$

Therefore, $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$.

Example If S is finite or countable (where the elements of S can be put into $1-1$ correspondence with a subset of the integers), then these technicalities really do not arise, for we we define for a given sample space S,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathcal{B}$ (why?).[hint: for each element, it is either in or out of a subset, so 2 choices].

Example Let $S = (-\infty, \infty)$, the real line. Then $\mathcal{B}$ is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathcal{B}$, it follows that $\mathcal{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

Definition Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function Pr with domain $\mathcal{B}$ that satisfies

1. $\Pr(A) \geqslant 0$ for all $A \in \mathcal{B}$.
2. $\Pr(S) = 1$.
3. If $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

Example Consider the simple experiment of tossing a fair coin (just once), so $S = \{H, T\}$. A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since $S = \{H\} \cup \{T\}$, we have , from Axiom 1, $\Pr(\{H\} \cup \{T\}) = 1$. Also, $\{H\}$ and $\{T\}$ are disjoint, so $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$. Collectively, we have

$$\begin{aligned}
&\Pr(\{H\}) = \Pr(\{T\}) \\
&\Pr(\{H\} \cup \{T\}) = 1 \\
&\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})
\end{aligned}$$

Therefore, $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$.

Caculus of Probabilities

We start with some fairly self-evident properties of the probability function when applied to a single event.

Theorem If Pr is a probability function and A is any set in $\mathcal{B}$, then

1. $\Pr(\varnothing) = 0$, where $\varnothing$ is the empty set;
2. $\Pr(A) \leqslant 1$;
3. $\Pr(A^c) = 1 - \Pr(A)$.

proof:

Theorem If Pr is a probability function and A and B are any sets in $\mathcal{B}$, then

1. $\Pr(B \cap A^c) = \Pr(B) - \Pr(A \cap B)$;
2. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$;
3. If $A \subset B$, then $\Pr(A) \leqslant \Pr(B)$.

proof:

Formula (2) in the above theorem gives a useful inequality for the probability of an intersection (Bonferroni's Inequality):

$$\Pr(A \cap B) \geqslant \Pr(A) + \Pr(B) - 1.$$

Theorem If Pr is a probability function, then

1. $\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \cap C_i)$ for any partition $C_1, C_2, \ldots$;
2. $\Pr(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{\infty} \Pr(A_i)$ for any sets $A_1, A_2, \ldots$.

where (1) is also referred to as "Total probability" and (2) is Boole's inequality.
proof:

Conditional Probability

All of the probabilities that we have dealt with thus far have been unconditional probabilities. A sample space was defined and all probabilities were calculated with respect to that sample space. In many instances, however, we are in a position to update the sample space based on new information. In such cases we want to be able to update probability calculations or to calculate *conditional probabilities*.

Definition If A and B are events in S, and $\Pr(B) > 0$, then the *conditional probability* of A given B, written $\Pr(A|B)$, is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Note that B becomes the sample space now: $\Pr(B|B) = 1$.

Example Four cards are dealt from the top of a well-shuffled deck. What is the probability that they are the four aces? (there are in total 52 cards)

solution:

Theorem (Bayes' Rule) Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots,$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^{\infty} \Pr(B|A_j)\Pr(A_j)}.$$

proof:

Independence

Definition Two events, A and B, are *statistically independent* if

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

Note that independence could have been defined using Bayes' rule by $\Pr(A|B) = \Pr(A)$ or $\Pr(B|A) = \Pr(B)$ as long as $\Pr(A) > 0$ or $\Pr(B) > 0$. More notation, often statisticians omit $\cap$ when writing intersection in a probability function which means $\Pr(AB) = \Pr(A \cap B)$. Sometime, statisticians use comma (,) to replace $\cap$ inside a probability function too, $\Pr(A, B) = \Pr(A \cap B)$.

Theorem If A and B are independent events, then the following pairs are also independent.

1. A and B^c,
2. A^c and B,
3. A^c and B^c.