Lab 1

Foundations of Probability Theory

Abstract: Probability theory is the foundation of statistical science, providing a mathematical means of modeling random experiments or uncertainty. Through these mathematical models, researchers are able to draw inferences about the random experiments using observed data. The aim of this lab is to outline the basic ideas of probability theory that are fundamental to the study of statistics. The entire structure of probability, and therefore of statistics, can be built on this relatively straightforward foundation.

Key words: Bayes' theorem, σ-field, Complement, Conditional probability, Event, Independence, Intersection, Multiplication rule, Probability function, Probability space, Random experiment, Sample space, Sets, Sigma algebra, Union.

1.1. Random Experiments

A random experiment is a mechanism which has at least two possible outcomes. When a random experiment is performed, one and only one outcome will occur, but which outcome to occur is unknown in advance. In other words, a random experiment is a mechanism for which the outcome cannot be predicted with certainty.

In statistics, the purpose of mathematical statistics is to provide mathematical models for random experiments of interest. Once a model for such an experiment has been provided and the theory worked out in detail, the statistician may, within this framework, make inference (i.e., draw conclusions) about the probability law of the random experiment based on observed data.

1.2. Basic Concepts of Probability

Definition 1.2.1 (Sample Space)**.** The possible outcomes of a random experiment are called "basic outcomes", and the set of all basic outcomes constitutes "the sample space", which is denoted by S. When an experiment is performed, the realization of the experiment is one (and only one) outcome in the sample space. If the experiment is performed a number of times, different outcomes may occur each time or some outcomes may repeat.

Example 1.2.1 (Throwing a coin)**.** Two possible outcomes: heads or tails. The sample space is $S = \{H, T\}$.

Example 1.2.2 (Rolling a Die)**.** The basic outcomes are the numbers 1, 2, 3, 4, 5, 6. The sample space $S = \{1, 2, 3, 4, 5, 6\}$.

Example 1.2.3 (Throwing two coins)**.** We have sample space

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

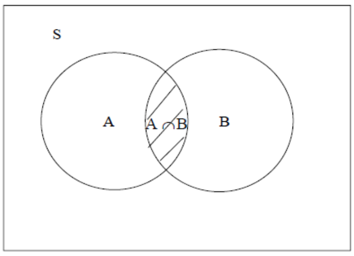
Definition 1.2.2 (Event)**.** An event A is a collection of basic outcomes from the sample space S that share certain common features or equivalently obey certain restrictions. The event A is said to occur if the random experiment gives rise to one (and only one) of the constituent basic outcomes in A. That is, an event occurs if any of its basic outcomes has occurred (or equivalently if the outcome of the random experiment is an element of event A).

Mathematically speaking, an event is equivalent to a set. Henceforth, the words "set" and "event" are interchangeable.

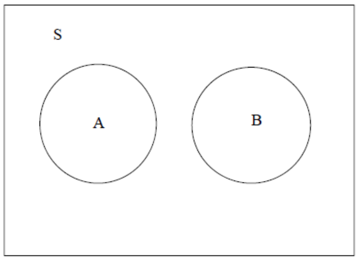
Example 1.2.4. Event A is defined as "the number resulting is even". Event B is "the number resulting is at least 4". Then $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$.

1.3. Set Theory

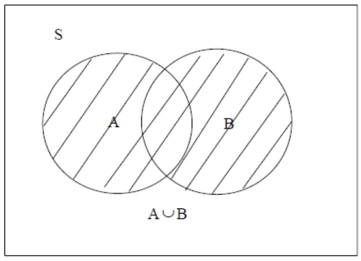
Probability theory builds upon set theory, or the algebra of sets, which we begin now. The Venn Diagram, originally introduced by Venn in his book, *Symbolic Logic*, published in 1881, can be used to depict a sample point, a sample space, an event, and related concepts. Specifically, the Venn Diagram is made up of two or more overlapping circles, where one circle denotes one set.



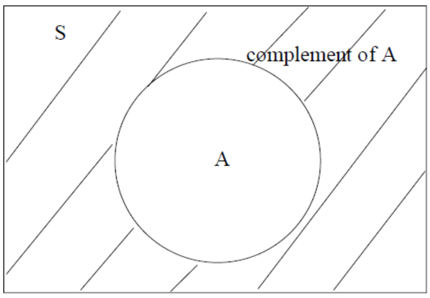
$A \cap B$



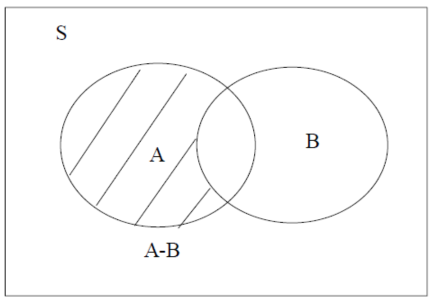
Two exclusive events



$A \cup B$



A^c



$A - B$

Theorem 1.3.1 (Laws of Set Operations)**.**
Complementation

$$(A^c)^c = A$$
$$\emptyset^c = S$$
$$S^c = \emptyset$$

Commutativity

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

Associativity

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

Distributivity

$$B \cap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (B \cap A_i)$$
$$B \cup \left(\bigcap_{i=1}^{n} A_i \right) = \bigcap_{i=1}^{n} (B \cup A_i)$$

De Morgan's Laws

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

Proof. All these proofs are based off the technique of "element choosing". For example, the proof for De Morgan's first law is

$$a \in \left(\bigcup_{i=1}^{n} A_i \right)^c \iff a \notin \bigcup_{i=1}^{n} A_i \iff a \notin A_i \quad \forall i \iff a \in A_i^c \quad \forall i \iff a \in \bigcap_{i=1}^{n} A_i^c$$

Rest are exercises. □

Example 1.3.1. Suppose A and B are disjoint. Under what conditions are A^c and B^c disjoint?

Solution. A^c and B^c are disjoint if and only if $A \cup B = S$. The proof is as follows:

$$A^c \text{ and } B^c \text{ disjoint} \overset{\text{definition}}{\iff} A^c \cap B^c = \emptyset \iff (A \cup B)^c = \emptyset \iff A \cup B = S$$

where the middle $\iff$ applies De Morgan's Laws. □

Example 1.3.2. Let A and B be two events in S. Then

- Are $A \cap B$ and $A^c \cap B$ mutually exclusive?
- Is $(A \cap B) \cup (A^c \cap B) = B$?
- Are A and $A^c \cap B$ mutually exclusive?
- $A \cup (A^c \cap B) = A \cap B$?

Solution. Answers are

- Yes. $(A \cap B) \cap (A^c \cap B) = (A \cap A^c) \cap (B \cap B) = \emptyset \cap B = \emptyset$, by associativity and commutativity.
- Yes. $(A \cap B) \cup (A^c \cap B) = (A \cup A^c) \cap B = S \cap B = B$, by distributivity.
- Yes. $A \cap (A^c \cap B) = (A \cap A^c) \cap B = \emptyset \cap B = \emptyset$, by associativityy.
- No. $A \cup (A^c \cap B) = (A \cup A^c) \cap (A \cup B) = S \cap (A \cup B) = A \cup B$.

□

1.4. Fundamental Probability Laws

Definition 1.4.1. A collection of subsets of S is called a σ-algebra (or Borel Field), denoted by $\mathcal{B}$, if it satisfies the following three properties:

- $\emptyset \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$)
- If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation)
- If $A_1, A_2, ... \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions)
  - Alternatively, if $A_1, A_2, ... \in \mathcal{B}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable intersections)

Definition 1.4.2. Given a sample space S and an associated sigma algebra $\mathcal{B}$, a probability function is a function P with domain $\mathcal{B}$ that satisfies

- $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{B}$
- $\mathbb{P}(S) = 1$
- If $A_1, A_2, ... \in \mathcal{B}$ are pairwise disjoint, then $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Theorem 1.4.1. If P is a probability function and A is any set in $\mathcal{B}$, then

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A) \leq 1$
- $\mathbb{P}(\emptyset) = 0$

Proof. We have

- Since $S = A \cup A^c$, and $A \cap A^c = \emptyset$, we have
$$1 = \mathbb{P}(S) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$$
- Suppose $\mathbb{P}(A) > 1$. Using the result just proved, this would imply $\mathbb{P}(A^c) < 0$, a contradiction to a probability function.
- Let $A = S$ and $A^c = \emptyset$.

□

Theorem 1.4.2. If P is a probability function and A and B are any sets in $\mathcal{B}$, then

- $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
  - Generalization:
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \le i < j \le n} \mathbb{P}(A_i \cap A_j)$$
$$+ \sum_{1 \le i < j < k \le n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$
- If $A \subseteq B$, then $\mathbb{P}(A) \le \mathbb{P}(B)$

Proof. We have

- Since $B = (A^c \cap B) \cup (A \cap B)$ and $(A^c \cap B) \cap (A \cap B) = \emptyset$, we have
$$\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B)$$
- Since $A \cup B = A \cup (A^c \cap B)$ and $A \cap (A^c \cap B) = \emptyset$, we have
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$$
  Using the result from before
$$\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B)$$
  Subtract the last two equations to get the result.
- If $A \subseteq B$, then $A \cap B = A$. Apply the previous result to have
$$\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(A^c \cap B) \ge 0$$

□

Theorem 1.4.3. If P is a probability function, then

- $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap C_i)$ for any partition $C_1, C_2, ...$
- $\mathbb{P}(\cup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for any sets $A_1, A_2, ...$ (Boole's Inequality)

Proof. Lecture notes. □

Example 1.4.1. Let $A_1, \cdots, A_4$ be events such that $\mathbb{P}(A_j) = 1/2$ for $j = 1, \cdots, 4$. Prove that there exists a joint event $A_j \cap A_k$ that occurs with at least probability $1/6$. More formally,
$$\max_{1 \le j < k \le 4} \mathbb{P}(A_j \cap A_k) \ge \frac{1}{6}$$

Solution. Using the generalized inclusion-exclusion principle,

$$
\begin{aligned}
1 &\geq P\left(\bigcup_{i=1}^{4} A_i\right) \\
&= \sum_{i=1}^{4} \mathbb{P}(A_i) - \sum_{1\leq i<j\leq 4} \mathbb{P}(A_i \cap A_j) + \underbrace{\sum_{1\leq i<j<k\leq 4} \mathbb{P}(A_i \cap A_j \cap A_k) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)}_{\geq 0} \\
&\geq 2 - \sum_{1\leq i<j\leq 4} \mathbb{P}(A_i \cap A_j)
\end{aligned}
$$

which gives us

$$\sum_{1\leq i<j\leq 4} \mathbb{P}(A_i \cap A_j) \geq 1 \implies \frac{1}{6} \sum_{1\leq i<j\leq 4} \mathbb{P}(A_i \cap A_j) \geq \frac{1}{6}$$

Since the average of the $\mathbb{P}(A_i \cap A_j)$ is at least 1/6, clearly the max is at least 1/6. □

1.5. Conditional Probability

Definition 1.5.1. If A and B are events in S, and $\mathbb{P}(B) > 0$, then the conditional probability of A given B, written $\mathbb{P}(A|B)$, is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Example 1.5.1. Let $A, B, C_1, \cdots, C_n$ be events on probability space $(S, \mathcal{B}, P)$. Suppose that $\bigcup_{i=1}^{n} C_i = S$ and $\mathbb{P}(C_i) > 0$, $\mathbb{P}(A|C_i) \geq \mathbb{P}(B|C_i)$ for any $i = 1, \cdots, n$. Is it true that $\mathbb{P}(A) \geq \mathbb{P}(B)$?

Solution. No. Let $\mathcal{B} = \{b_1, b_2, b_3\}$ and $A = \{b_2\}, B = \{b_1, b_3\}, C_1 = \{b_1, b_2\}, C_2 = \{b_2, b_3\}$, and let P be the uniform measure. *Remark:* The answer would be 'yes' if we also require the C_i's be disjoint. □

Example 1.5.2. A hotel packed breakfast for each of three guests. There are three types of rolls (nut, cheese, fruit). The preparer wrapped each of the nine rolls and randomly put three rolls in a bag for each of the guests. What is the probability that guests received one roll of each type?

Solution. Let A_1, A_2, A_3 denote the events that the rolls sequentially put in Bag 1 are all different. Let B_1, B_2, B_3 denote the events that the rolls sequentially put in Bag 2 are all different, and similarly for $C_1, \cdots, C_3$ for Bag 3. Let $A = \{A_1, A_2, A_3\}$, and similarly for B and C. Then,

$$
\begin{aligned}
\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_3|A_2, A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_1) &= \frac{3}{7} \cdot \frac{6}{8} \cdot \frac{9}{9} = \frac{9}{28} \\
\mathbb{P}(B_1 \cap B_2 \cap B_3|A) = \mathbb{P}(B_3|B_2, B_1, A)\mathbb{P}(B_2|B_1, A)\mathbb{P}(B_1|A) &= \frac{2}{4} \cdot \frac{4}{5} \cdot \frac{6}{6} = \frac{2}{5} \\
\mathbb{P}(C_1 \cap C_2 \cap C_3|A, B) = \mathbb{P}(C_3|C_2, C_1, A, B)\mathbb{P}(C_2|C_1, A, B)\mathbb{P}(C_1|A, B) &= \frac{1}{1} \cdot \frac{2}{2} \cdot \frac{3}{3} = 1
\end{aligned}
$$

Hence $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap B_1 \cap B_2 \cap B_3 \cap C_1 \cap C_2 \cap C_3) = \frac{9}{28} \cdot \frac{2}{5} \cdot 1 = \frac{9}{70}$. □

1.6. Bayes' Theorem

Theorem 1.6.1. Let $A_1, A_2, ...$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, ...$,

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

Example 1.6.1. Suppose a disease has prevalence $\alpha \in (0, 1)$ throughout a population. Suppose a test to detect the disease has sensitivity (true positive rate) of θ and specificity ϕ (true negative rate). Having taken BIO230, you know that such a test can be wildly inaccurate, so you take the test K times just to be sure. All K tests reveal that you do have the disease. Show that,

$$\lim_{K\to\infty} \mathbb{P}(\text{Truly } + |\text{All } K \text{ tests positive}) = 1$$

if and only if $\theta + \phi > 1$.

Solution. By Bayes' rule,

$$\begin{aligned}\mathbb{P}(\text{Truly } + |\text{All } K \text{ tests positive}) &= \frac{\mathbb{P}(\text{All } K \text{ tests positive}|\text{Truly } +)\mathbb{P}(\text{Truly } +)}{\mathbb{P}(\text{All } K \text{ tests positive})} \\ &= \frac{\theta^K \alpha}{\theta^K \alpha + (1-\phi)^K (1-\alpha)} \\ &= \frac{\alpha}{\alpha + (\frac{1-\phi}{\theta})^K (1-\alpha)}\end{aligned}$$

Notice that $\lim_{K\to\infty} \mathbb{P}(\text{Truly } + |\text{All } K \text{ tests positive}) = 1$ if and only if $\frac{1-\phi}{\theta} < 1$, hence the result follows. □

1.7. Independence

Definition 1.7.1. Two events, A and B are statistically independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Definition 1.7.2. A collection of events $A_1, ..., A_n$ are mutually independent if for any subcollection $A_{i_1}, ..., A_{i_k}$, we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

Example 1.7.1. Suppose A, B are independent. Is it true that $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$ for any arbitrary, non-null event C? Prove or disprove.

Solution. Not true. Suppose we flip two fair coins. Let A be the event the first is heads, and B be the even the second is heads, and C be the event that the two coins are the same. Clearly A, B are independent, but

$$\mathbb{P}(A \cap B|C) = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

□

1.8. Additional Problems

1. We have six ropes. Each rope has its ends labeled A and B. We randomly select the A ends, two at a time, and tie them together. Perform this similarly for the B ends. What is the probability that we have a single ring after performing this?

2. Container A initially contains 1000 green and 3000 red apples and container B contains 3000 green and 1000 red apples. We randomly choose 2000 apples from container A and place them in container B. One apple is then chosen at random from container B. What is the probability that it is green?

3. Archers A and B are successively shooting at a target, with A going first. Suppose the probability A hits the target is p_A, and the probability B hits is p_B. The first to hit the target wins. What is the probability A wins?

4. Give an example of 4 dependent random events $A_1, \cdots, A_4$ such that any 3 of them are independent.

5. Let $A_1, \cdots, A_{2000} \subseteq A$ with $|A_i| \geq 6$ and at least one pair i, j such that $A_i \neq A_j$ (that is, not all A_i are the same). Prove that there are more than 100 distinct partitions of A into 5 disjoint subsets $E_1, \cdots, E_5$ such that each A_i contains elements that belong to at least two of the E_i.

6. A function $d : X \times X \to \mathbb{R}$ is called a *metric* if it satisfies

   - $d(x_1, x_2) \geq 0$ with equality if and only if $x_1 = x_2$
   - $d(x_1, x_2) = d(x_2, x_1)$ for all $x_1, x_2 \in X$
   - $d(x_1, x_2) \leq d(x_1, x_3) + d(x_2, x_3)$ for all $x_1, x_2, x_3 \in X$

   Define $d_1, d_2 : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ as

$$d_1(A, B) = \mathbb{P}(A \Delta B)$$
$$d_2(A, B) = \begin{cases} \frac{\mathbb{P}(A \Delta B)}{\mathbb{P}(A \cup B)}, & \text{if } \mathbb{P}(A \cup B) \neq 0 \\ 0, & \text{if } \mathbb{P}(A \cup B) = 0 \end{cases}$$

   where $A \Delta B = (A \cup B) - (A \cap B) = (A^c \cap B) \cup (A \cap B^c)$ (symmetric difference). Prove that d_1, d_2 are valid metrics.