Theory Notes

Note: Finished Lecture 5

Introduction

- **Probability** is a branch of mathematics concerned with the study of *random phenomenon* (e.g., experiments, models of populations).
- We are primarily interested in probability as it relates to **statistical inference**, the science of drawing inferences about populations based on only a part of the population (i.e., a sample).

Some Definitions

1. **population**: the entire set of objects that we are interested in studying
   e.g., all ISU students
2. **sample**: the subset of the population available for observation
   e.g., STAT 542 students

Note: population and sample are crucial terms in understanding statistics (i.e., STAT 543), but will not occur very often in our discussions of probability theory (i.e., STAT 542).

3. **experiment**: process of obtaining an observed result of a random phenomenon
4. **sample space** S: the set of all possible outcomes of the experiment

- elements $s \in S$ of a sample space are called **sample points** (s)
- a sample space may be
- **discrete**
  (finite or countably infinite, i.e., listable as a finite/infinite sequence)

$$S = \{s_1, s_2, \ldots, s_n\}$$

or

$$S = \{s_1, s_2, s_3, \ldots\}$$

- or **continuous**
  (uncountably infinite, i.e., a continuum of sample points like
  $S = [0, \infty)$)

5. **event** (e.g., $A, B, \ldots$): subset of the sample space S

- **set**: A is a collection of elements
  (in our case, A is a collection of outcomes)
- **membership**: $x \in A$ or $x \notin A$
  (x is in A or x is not in A)
- **complement**:
  $$A^c = \{x : x \notin A\}$$
  (x such that x is not in A)
- **union**:
  $$A \cup B = \{x : x \in A \text{ or } x \in B\}$$
  (x is in A or B or both)
- **intersection**:
  $$A \cap B = \{x : x \in A \text{ and } x \in B\}$$
- **subset**: $A \subset B$ means that A is contained in B
  (formally, $x \in A \Rightarrow x \in B$)
- **equality**: $A = B$ if $A \subset B$ and $B \subset A$
- **empty set**: $\varnothing$

Algebraic Laws

- **commutativity**:
  $$A \cup B = B \cup A$$
  $$A \cap B = B \cap A$$
- **associativity**:
  $$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$$
  $$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$$
- **distributive law**:
  $$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
  $$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
- **DeMorgan's laws**:
  $$(A \cup B)^c = A^c \cap B^c$$
  $$(A \cap B)^c = A^c \cup B^c$$

Aside on disjoint and partitions

- events A and B are **disjoint** (mutually exclusive) if

$$A \cap B = \varnothing$$

- For a sequence $A_1, A_2, \ldots$ of events, we say $A_1, A_2, \ldots$ are **pairwise disjoint** if

$$A_i \cap A_j = \varnothing \quad \text{for all } i \neq j$$

- $A_1, A_2, \ldots$ is a **partition** of S if the A_i's are pairwise disjoint and exhaustive, that is,

$$\bigcup_{i=1}^{\infty} A_i = S \quad \text{and} \quad A_i \cap A_j = \varnothing \quad \text{for all } i \neq j$$

Probability Functions

- A **probability function** is a function P defined on a Borel field $\mathcal{B}$ of the sample space S that satisfies:
  1. $P(A) \geq 0$ for all $A \in \mathcal{B}$
  2. $P(S) = 1$
  3. If $A_1, A_2, \ldots \in \mathcal{B}$ are *pairwise disjoint*, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- Any function satisfying the above is a legitimate probability function.

Theorem 1.2.8.
If P is a probability function and A is any set in $\mathcal{B}$, then:

(a)

$$P(\varnothing) = 0$$

(b)

$$P(A) \leq 1$$

(c)

$$P(A^c) = 1 - P(A)$$

Proof of (c) (parts (a) and (b) follow from (c) and the axioms):

Since

$$S = A \cup A^c,$$

and A and A^c are disjoint, by the axioms of probability,

$$P(S) = P(A \cup A^c) = P(A) + P(A^c).$$

Because $P(S) = 1$, we have

$$1 = P(A) + P(A^c),$$

which implies

$$P(A^c) = 1 - P(A).$$

Theorem 1.2.9.
If P is a probability function and A, B are sets in $\mathcal{B}$, then:

(a)

$$P(B \cap A^c) = P(B) - P(B \cap A)$$

(b)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(c) If $A \subset B$, then

$$P(A) \leq P(B).$$

Theorem 1.2.11.
If P is a probability function, then

(a) For any partition $C_1, C_2, \ldots \in \mathcal{B}$ (i.e., disjoint C_i's and $\bigcup_{i=1}^{\infty} C_i = S$),

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

(b) For any sets $A_1, A_2, \ldots \in \mathcal{B}$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

Principle of Inclusion–Exclusion.
For any sets $A_1, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

Equivalently,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

This generalizes

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

and is proven by induction.

Bonferroni's Inequalities.
For any sets $A_1, \ldots, A_n$ and any $m \in \{1, \ldots, n\}$,

- if m is odd,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right),$$

- if m is even,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \ge \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

In particular,

$$\sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) \le P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i).$$

Combinatorics

Permutations / ordered arrangements II.
When selecting r objects from n objects (without replacement), the number of ordered arrangements possible is

$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Combinations / unordered selections.
The number of ways to choose r objects from n objects (without replacement), where the ordering doesn't matter, is

$$\binom{n}{r} \equiv \frac{n!}{r!\,(n-r)!}.$$

Summary table: number of ways to select r objects from a group of n

	objects chosen **without replacement**	objects chosen **with replacement**
ordered	$\dfrac{n!}{(n-r)!}$	n^r
unordered	$\dbinom{n}{r}$	$\dbinom{n+r-1}{r}$

Conditional Probability

- **Definition:** If A, B are events in S with $P(B) > 0$, then
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$
- In conditioning, B can be thought of as the **updated sample space**, i.e., not all of S is relevant since we know B has occurred.

$P(\cdot \mid B)$ is a probability function that satisfies the usual axioms and properties.

Axioms:

- $P(A \mid B) \geq 0$ for all events A
- $P(B \mid B) = 1$
  (B is the updated sample space)
- If $A_1, A_2, \ldots$ are pairwise disjoint events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

Some properties:

$$P(A^c \mid B) = 1 - P(A \mid B)$$

$$P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 \cap A_2 \mid B)$$

It also follows from our definition of conditional probability that

$$P(A \cap B) = P(B \mid A)\, P(A) = P(A \mid B)\, P(B).$$

More generally, for events $A_1, A_2, \ldots, A_n$,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

It is possible to reverse the conditioning of A and B to obtain **Bayes' rule**:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

More generally, if $A_1, A_2, \ldots$ is a partition of the sample space S, then we obtain a general version of Bayes' rule:

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)\, P(A_j)}.$$

Independence

If $P(A \mid B) = P(A)$, then the occurrence of B does not affect the probability of A. It then follows that

$$P(A \cap B) = P(A)P(B) \quad \text{and} \quad P(B \mid A) = P(B).$$

We define two events A and B as **independent** if

$$P(A \cap B) = P(A)P(B).$$

More than two events.
$A_1, \ldots, A_n$ are **independent** if and only if, for any subcollection $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ of distinct indices (with any $2 \leq k \leq n$), it holds that

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

- If $A_1, \ldots, A_n$ are independent, then

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{for any } i \neq j.$$

- However,

$$P(A_i \cap A_j) = P(A_i)P(A_j) \text{ for } i \neq j$$

does **not** imply that $A_1, \ldots, A_n$ are independent.

If $A_1, \ldots, A_n$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n).$$

However,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

holding does **not** imply that $A_1, \ldots, A_n$ are independent.

The assumption of independence of events allows the computation of joint occurrences of events through simple calculations.

Random Variables

Definition: A **random variable** (r.v.) X is a function defined on a sample space S that associates a real number with each outcome in S.

That is, for each $s \in S$, we have

$$X(s) \in \mathbb{R}.$$

In function notation,

$$X : S \to \mathbb{R}.$$

We usually suppress the dependence of X on $s \in S$ and write

$$X = X(s).$$

We have $P(A)$ defined on events $A \subset S$, which can be used to assign probabilities for events concerning a random variable X on $\mathbb{R}$ ($X : S \to \mathbb{R}$).

Define $P_X(\cdot)$ for events $B \subset \mathbb{R}$ as follows:

$$P_X(B) = P_X(X \in B) = P\big(\{s \in S : X(s) \in B\}\big).$$

$P_X(\cdot)$ satisfies the axioms and is therefore a legitimate probability function.

CDF

Definition.
The **cumulative distribution function** (cdf) of a random variable X, denoted by $F(\cdot)$, is defined by

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Sometimes written with subscript as $F_X(x)$.

A function $F(x)$, $x \in \mathbb{R}$, is a cdf for some random variable if and only if the following hold:

1. $F(x)$ is a nondecreasing function of x.
2. $$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$
3. $F(x)$ is right continuous, i.e.,
$$\lim_{x \downarrow x_0} F(x) = F(x_0) \quad \text{for any } x_0 \in \mathbb{R}.$$