# Syllabus

- Meeting Time: Tue, Thr 9:00 - 10:15am
- Pre-recommended: Mathematical Statistics and Regression Analysis (Undergraduate level)
- TEXTBOOK: Statistical Inference, 2nd edition; Casella and Berger, 2002.
- Homework 10%, Test I & II, 20% each, Midterm & Final 25%
- HOMEWORK: Approximately 10 homework assignments
- All exams are closed book and notes.

The purpose of this chapter is to define probability and discuss some of its properties. Probability is the foundation upon which almost all statistics is built, and provides a means for modeling populations, experiments, sampling and almost anything that could be considered a random phenomenon. Just as statistics builds upon the foundation of probability theory, probability theory builds on set theory

# 1.1. Set Theory

- Probability of "SOMETHING"
- Basics
  - **Set:** A collection of objects (or elements). Use the uppercase letters for set $(A, B, C, \cdots)$ and the lowercase letters for the elements $(x, y, z \cdots)$.
  - **Family or Class:** Set of sets. Elements are sets.
  - **Notations:**
    Element: $x \in A$, $x \notin A$
    Subset: $A \subset B$, $x \in A \rightarrow x \in B$
    Equivalent sets: $A = B$, $A \subset B$ and $B \subset A$
    Empty set: $\emptyset$, no element in the set
    Universal set: $\Omega$

# 1.1. Set Theory

### Definition (Sample Space $S$)

The set, $S$, of all possible outcomes of a particular experiment.

Example:

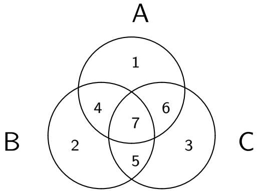
| Experiment | Sample Space |
|---|---|
| Tossing a coin | $S = \{H, T\}$ |
| Measuring weights | $S = \{x : 0 \leq x \leq 500\}$ |

### Definition (Event)

Any collection of possible outcomes of an experiment. That is, any subset of sample space $S$.

# 1.1. Set Theory

- Elementary Set Operations
  - Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$.
  - Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
  - Compliment: $A^c = \{x : x \notin A\}$.
  - Set difference:
    $A \backslash B = A - B = \{x : x \in A \text{ and } x \notin B\} = A \cap B^c$.
- Venn Diagram



# 1.1. Set Theory

## Theorem

**Idempotent Law:**

$A \cap A = A$, $A \cup A = A$.

**Commutative Law:**

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

**Distributive Law:**

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

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

**DeMorgan's Law:**

$(A \cap B)^c = A^c \cup B^c \quad , \quad (A \cup B)^c = A^c \cap B^c$

Note: Set Operations of more than two sets

$$\cup_{i=1}^{\infty} A_i \ , \ \cap_{i=1}^{\infty} A_i \ , \ \cup_{a \in \Gamma} A_a \ , \ \cap_{a \in \Gamma} A_a \quad .$$

# 1.1. Set Theory

### Definition

Two events $A$ and $B$ are *mutually exclusive* if $A \cap B = \emptyset$.

### Definition

$\{A_1, A_2, \cdots\}$ is called a *partition* of a sample space if $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $A_1 \cup A_2 \cup \cdots = S$.

# 1.2. Probability Theory

Probability:

- Frequency of occurrence:
  Experiment is performed a number of times. Check the frequency of outcomes of interest.
  Probability $=$ number of outcomes of interest/ number of experiments
- Axiomatic Probability:
  Set function that satisfies a certain axioms

## Definition (Sigma Algebra, or $\sigma$-algebra, or $\sigma$-field, or Borel Field $\mathcal{B}$)

A collection of subsets of $S$, $\mathcal{B}$, that satisfies
S1: $\emptyset \in \mathcal{B}$
S2: If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$
S3: If $A_1, A_2, \cdots \in \mathcal{B}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$

Note: S1-S3 also implies $S \in \mathcal{B}$ and it is closed under countable intersection.
Example: Let $\mathcal{B}$ be a set of all subsets of $\{1, 2, 3\}$. That is

$$\mathcal{B} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Is this sigma algebra ? If $S$ is finite or countable, then

$$\mathcal{B} = \{\text{all subset of } S, \ \text{including} S\}$$

# 1.2. Probability Theory

## 1.2.1. Axiomatic Foundation (Axioms of Probability)

Axioms of Probability

### Definition

Given $S$ and associated sigma algebra $\mathcal{B}$, a *probability function* is a function $P$ with domain $\mathcal{B}$ that satisfies

A1: $P(A) \geq 0$ for all $A \in \mathcal{B}$

A2: $P(S) = 1$

A3: If $A_1, A_2, \dots \in \mathcal{B}$ and pairwise mutually exclusive then

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

- Associated sigma algebra
- Probability Space: $(S, \mathcal{B}, P)$.

- Examples

1. Three-sided die with number 1, 2 and 3. $S = \{1, 2, 3\}$

$$\mathcal{B} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

Define $P$ as
$P(\emptyset) = 0$
$P(\{1\}) = P(\{2\}) = P(\{3\}) = 1/3$
$P(\{1,2\}) = P(\{1,3\}) = P(\{2,3\}) = 2/3$
$P(\{1,2,3\}) = P(S) = 1$

2. $S = [0, \infty)$, $\mathcal{B} =$?. Define $P$ as, for $A \in \mathcal{B}$, $P(A) = \int_A e^{-x} dx$ .
(See Example 1.2.3 for $S = (-\infty, \infty)$)

### Theorem

*(See Theorem 1.2.8.) Given* $(S, \mathcal{B}, P)$, *for* $A \in \mathcal{B}$,
*1)* $P(\emptyset) = 0$
*2)* $P(A) \leq 1$
*3)* $P(A^c) = 1 - P(A)$

### Theorem

*(See Theorem 1.2.9.) Given* $(S, \mathcal{B}, P)$, *for* $A, B \in \mathcal{B}$,
*1)* $P(B \cap A^c) = P(B) - P(A \cap B)$
*2)* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
*3) If* $A \subset B$ *then* $P(A) \leq P(B)$
*4)* $P(A \cap B) \geq P(A) + P(B) - 1$

4) Bonferroni's inequality

Consider the sample space with finite number of elements.

$$S = \{s_1, s_2, \cdots, s_N\} \ .$$

Let all outcomes(elements) are equally likely, that is, $P(\{s_i\}) = 1/N$, for all $i$. Then, for $A \subset S$,

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S}$$

satisfies 3 axioms of probability.

### Theorem (Fundamental Theorem of Counting)

*If a job consist of $k$ separate tasks, the $i$-th of which can be done in $n_i$ ways, $i = 1, \cdots, k$, then the entire job can be done in $n_1 \times \cdots \times n_k$ ways.*

### Definition

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$$

### Definition (Binomial Coefficient)

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad , \quad n \geq r$$

# 1.3. Conditional Probability & Independence

• Probability that we have rain tomorrow: Probability of rain tomorrow given rain today.

• Conditional probability is a probability defined on an updated sample space based on the available information.

▷ Example: Toss a fair die. $A = \{1\}$, $B = \{1, 3, 5\}$. What i s the probability of throwing a 1 given that an odd numbers is thrown ?

## Definition

If $A$ and $B$ are events in $S$, and $P(B) > 0$, then the *conditional probability* of $A$ given $B$ is

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

# 1.3. Conditional Probability & Independence

- Note: Conditional probability also satisfies 3 axioms. That is,

A1: $P(A|B) \geq 0$ for all $A \in \mathcal{B}$
A2: $P(S|B) = 1$
A3: If $A_1, A_2, \cdots \in \mathcal{B}$ and pairwise mutually exclusive then

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

$\triangleright$ Example: Medical Test
In case of disease: test will fail to detect on 10% of cases
In case of no-disease: test will produce a false positive in 20% of cases.
We have 30% of individuals tested actually have disease.

# 1.3. Conditional Probability & Independence

▷ Example: Medical Test - Continued

Want to know (i) lab's overall error rate and (ii) the likelihood of having disease if lab return a positive result (+).

$$P(+|D) = 0.9, \qquad P(-|D) = 0.1,$$

$$P(+|N.D.) = 0.2, \qquad P(-|N.D.) = 0.8, \qquad P(D) = 0.3.$$

| Prob | Positive | Negative |
|---|---|---|
| Disease | 0.27(=0.9·0.3) | 0.1×0.3 |
| No Disease | 0.2×0.7 | 0.8×0.7 |

$P(\text{overall error}) = 0.03 + 0.14$

$P(\text{Disease} \mid +) = 27/41$ by Bayes' rule.

# 1.3. Conditional Probability & Independence

## Theorem

*Let* $\{A_1, A_2 \cdots\}$ *be a partition of the sample space* $S$ *and* $B$ *be any subset of* $\mathcal{B}$. *Then*

$$P(B) = \sum_{j=1}^{\infty} P(A_j \cap B) = \sum_{j=1}^{\infty} P(B|A_j)P(A_j) \quad ,$$

*(law of the total probability)*
*and for each* $i = 1, 2, \cdots$

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

*(Bayes' Rule)*

# 1.3. Conditional Probability & Independence

$\triangleright$ Example:

There are three rental agencies $A$, $B$ and $C$. Probability of having unsafe car from these agencies are 0.1, 0.08 and 0.125 respectively. An agency is chosen randomly and a car tested is found to be unsafe. What is the conditional probability that the car came from agency $A$, $B$ or $C$ ?

$P(Unsafe) =$

$P(A|Unsafe) =$

$P(B|Unsafe) =$

$P(C|Unsafe) =$

# 1.3. Conditional Probability & Independence

### Definition

Events $A_1, A_2, \cdots, A_n$ are called *mutually independent* provided the probability of the intersection of any sub-collection of events is the product of the probabilities of the events in the sub-collection. That is, if for any sub-collection $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$,

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

# 1.4. Random Variables

In most cases, we have interest in probabilities of certain event or function of elements of the sample space.

$$(S, \mathcal{B}, P) \overset{R.V.\ (X)}{\Longrightarrow} (\mathcal{R}, \mathcal{B}^1, P_X)$$

$\mathcal{R}$ : Real Number , $\mathcal{B}^1$ : $\sigma$-field generated by $R$

$P_X$ : Induced probability from $P$

## Definition

A random variable is a function from a sample space $S$ into the real numbers.

# 1.4. Random Variables

▷ Example: Toss a fair coin 3 times

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

$$\mathcal{B} = \text{Collection of all subsets of } S$$

Let $X$ be a number of heads. Then

| $s$ | $P(s)$ | $X(s) = x$ |
|---|---|---|
| hhh | 1/8 | 3 |
| hht | 1/8 | 2 |
| hth | 1/8 | 2 |
| htt | 1/8 | 1 |
| thh | 1/8 | 2 |
| tht | 1/8 | 1 |
| tth | 1/8 | 1 |
| ttt | 1/8 | 0 |

# 1.4. Random Variables

▷ Example - Continued

| $X = x$ | $P(X = x) = P_X(x)$ |
|---|---|
| 0 | |
| 1 | |
| 2 | |
| 3 | |

$$P(X = 1) = P_X(1) = P\left(s \in S : X(s) = 1\right) = \frac{3}{8}$$

◁ Note:

Random variable: Uppercase letter $(X, Y, \cdots)$
Values of RV: Lowercase letter $(x, y, \cdots)$
Subscript $X$ is often deleted

# 1.5. Distribution Functions

Most general tool to specify the distribution (probability structure) of a random variable.

## Definition

The cumulative distribution function (cdf) for a random variable $X$ is

$$F_X(x) = P(X \leq x) \quad , \quad \text{for all } x \in \mathcal{R}$$

$\triangleleft$ Note:
$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$.

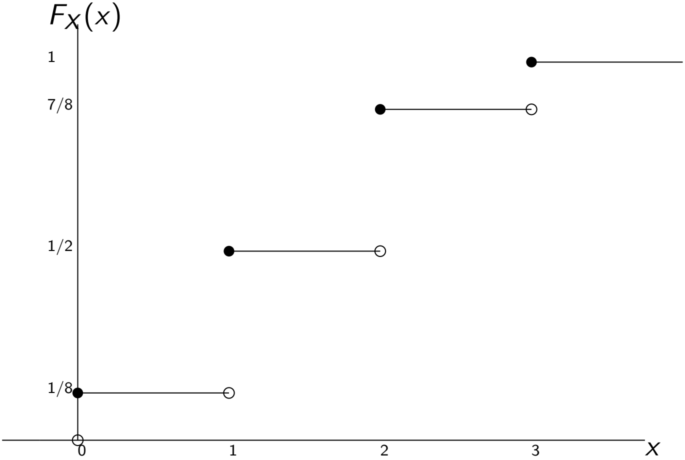
# 1.5. Distribution Functions

▷ Example: Tossing 3 coins. $X$= number of heads

$$F_X(x) = \begin{cases} \quad , & x < 0, \\ \quad , & 0 \le x < 1, \\ \quad , & 1 \le x < 2, \\ \quad , & 2 \le x < 3, \\ \quad , & 3 \le x \quad . \end{cases}$$

Figure of the cdf of this RV $X$: Step function.

# 1.5. Distribution Functions



# 1.5. Distribution Functions

## Theorem

*The function $F_X(x)$ is a cdf if and only if the following 3 conditions hold:*

1. $\lim_{x\to-\infty} F_X(x) = 0$ *and* $\lim_{x\to+\infty} F_X(x) = 1$
2. $F_X(x)$ *is nondecreasing*
3. $F_X(x)$ *is right continuous, that is,*

$$\lim_{x \downarrow x_0} F_X(x) = F_X(x_0) \quad , \quad \text{for all} \quad x_0 \in \mathcal{R} \quad .$$

## Definition

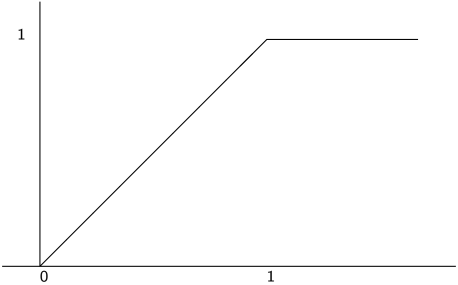
A random variable $X$ is *continuous* if $F_X(x)$ is a (absolute) continuous function of $x$. A random variable $X$ is *discrete* if $F_X(x)$ is a step function of $x$.

$\triangleright$ Example: Geometric distribution (See Example 1.5.4)
Free draw success percentage of a basketball player $= p$. Under the independence assumption from shot to shot, let $X$ be the number of shots required to get a hit. The *support* (collection of real numbers at which probability is positive) of $X$ is $\{1, 2, 3, \cdots\}$.
$P(X = x) = P(\{MMM \cdots H\}) =$
Thus, $F_X(x) =$

# 1.5. Distribution Functions

▷ Example: Uniform distribution on $[0, 1]$, $X \sim (0, 1)$

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \le x \le 1, \\ 1, & 1 < x\ . \end{cases}$$



# 1.5. Distribution Functions

## Lemma

*If the random variable $X$ is continuous with cdf $F_X(x)$ then $P(X = a) = 0$ for all real number $a$*

Proof) As $\{X = a\} \subset \{a - \epsilon < X \leq a\}$ for any $\epsilon > 0$, we have

$$
\begin{aligned}
P(X = a) \leq & P(a - \epsilon < X \leq a) \\
= & P(X \leq a) - P(X \leq a - \epsilon) \\
= & F_x(a) - F_x(a - \epsilon)
\end{aligned}
$$

Then, $0 \leq P(X = a) \leq \lim_{\epsilon \downarrow 0}[F_x(a) - F_x(a - \epsilon)] = 0$. as $F_x$ is continuous. Thus $P(X = a) = 0$.

# 1.5. Distribution Functions

▷ Example - Distribution of waiting time: Exponential distribution, $X \sim exp(\lambda)$

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-\lambda x), & 0 \leq x \end{cases},$$

for some $\lambda > 0$

◁ Note:

1. Probability at least $t_1$ min
2. Memoryless property: probability of waiting additional $t_1$ min after waiting $t_0$ min.

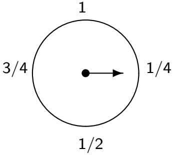
# 1.5. Distribution Functions

▷ Q - Is there a distribution that is neither discrete nor continuous? (Example 1.5.6)

1.

$$F_Y(y) = \begin{cases} \frac{1-\epsilon}{1+e^{-y}}, & y < 0, \\ \epsilon + \frac{1-\epsilon}{1+e^{-y}}, & y \geq 0 \; . \end{cases}$$

2. Consider an experiment composed of 2 steps. At the first step, a fair coin tossed. If we tail then define a random variable $X = 0$. If we have head, spin a fair spinnes marked (0,1] and let $X$ =ending position of spinnes.



# 1.5. Distribution Functions

2. - Continued

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1/2, & x = 0, \\ (x+1)/2, & 0 < x < 1, \\ 1, & 1 \leq x. \end{cases}$$

## Definition

The random variables $X$ and $Y$ are *identically distributed* if for every set $A \in \mathcal{B}^1$,

$$P(X \in A) = P(Y \in A)$$

Note that this does not imply $X = Y$.

# 1.6. Density & Mass Functions

The cdf is the most general, but not necessarily most appealing mean of specifying a probability on $\mathcal{R}$. Now let us consider the probability concerned with *point probability*.

### Definition

The *probability mass function* (pmf) associated with a discrete distribution with cdf $F_X$ is the function

$$f_X(x) = F_X(x) - \lim_{y \to x^-} F_X(y), \quad \text{(size of jump at point } x\text{)}$$

or

$$f_X(x) = P(X = x) \quad , \quad \text{for all} \quad x \in \mathcal{R} \; .$$

# 1.6. Density & Mass Functions

▷ Example: A four sided die that has different numbers (1, 2, 3, 4) affixed to each side. On a given roll each of 4 number is equally likely to occur. Roll the die twice. Define $X=$ maximum of two numbers.
pmf of $X$ is then

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $f(x)$ | $1/16$ | $3/16$ | $5/16$ | $7/16$ |

$$P(1 \leq X \leq 3) = P(X \leq 3) =$$

# 1.6. Density & Mass Functions

## Definition

The *probability density function* (pdf), $f_X(x)$, of a (absolute) continuous random variable with (absolute) continuous cdf $F_X$ is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } x \in \mathcal{R} \ .$$

Note: By the fundamental theorem of calculus,

$$\frac{dF_X(x)}{dx} = f_X(x).$$

▷ Example: Uniform distribution

# 1.6. Density & Mass Functions

**Theorem**

*A function $f_X(x)$ is a pdf (or pmf) of a random variable $X$ if and only if (iff)*

1. $f_X(x) \geq 0$ *for all* $x$
2. $\sum_x f_X(x) = 1$ *(pmf),* $\int_x f_X(x) = 1$ *(pdf)*

Note: The probability of $A \in \mathcal{R}$, then, is

$$P(X \in A) = \begin{cases} \sum_{x \in A} f_X(x), & \text{if discrete} \\ \int_A f_X(x), & \text{if continuous} \end{cases}$$

$\triangleright$ Example: A 12 sided die that has different numbers (1, 2, $\cdots$, 12) affixed to each side. On a given roll each of 12 number is equally likely to occur. Roll the die twice. Define $X=$ maximum of two numbers.
pmf of $X$ is then

| $x$ | 1 | 2 | $\cdots$ | 12 |
|---|---|---|---|---|
| $f(x)$ | | | | |

$$P(X = x) = c(2x - 1) \quad , \quad x = 1, 2, \cdots, 12 \quad .$$

Find a value for $c$.