# STA2001 Probability and Statistics (I)

## Lecture 5

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

# Review

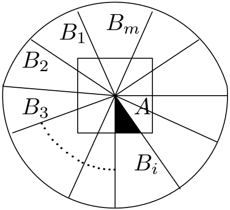
(Mutually) Independent Events:

- $A$ and $B$ are independent, if and only if any pair of the following events are independent
  (a) A and B'
  (b) A' and B
  (c) A' and B'
- $A, B, C$ are independent, if
  1. pairwise independent
  2. $P(A \cap B \cap C) = P(A)P(B)P(C)$

  Many properties hold.

# Review

Bayes' Theorem




Assume

1. $S = B_1 \cup B_2 \cup \cdots B_m, \quad B_i \cap B_j = \Omega$
2. $P(B_i) > 0$

Then

$$P(A) = \sum_{k=1}^{m} P(A \cap B_i) = \sum_{k=1}^{m} P(B_i) P(A|B_i)$$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)}, \text{ provided } P(A) > 0$$

# Chapter 2 Discrete Distribution

# Section 2.1 Random Variable of the Discrete Type

# Motivations

1. Flip a coin.
2. Select a color from 256 colors.

| original sample space | | new and numeric sample space |
|---|---|---|
| $S$ | $\leftrightarrow$ | $\overline{S}$ |
| $S = \{H, T\}$ | $\leftrightarrow$ | $\{1, 0\}$ |
| $S = \{R, G, \cdots, B\}$ | $\leftrightarrow$ | $\{1, 2, \cdots, 256\}$ |
| nonnumeric | | numeric |

There are other motivations ...

# Random Variable (RV)

**Definition[Random Variable]**

Given a random experiment with sample space $S$, a function $X : S \to \overline{S} \subseteq R$ that assign one real number $X(s) = x$ to each $s \in S$ is called a Random Variable (RV).

- $\overline{S}$ denote the range of $X$: $\overline{S} = \{x | X(s) = x, s \in S\}$.

# Understand a RV

**Question**

What's the relation between $S$ and $X$? What's the relation between $S$ and $\overline{S}$?

$$X : S \to \overline{S}$$

- RV defines a new random experiment with a numeric sample space $\overline{S}$
- If $X$ is one to one, then old random experiment with $S$

  $\Leftrightarrow$ new random experiment with $\overline{S}$
- If $X$ is not one to one, then old random experiment with $S$

  $\not\Leftrightarrow$ new random experiment with $\overline{S}$ (example will be given later)
- repeat the new random experiment is to generate a number randomly from $\overline{S}$

# Example 1

The old random experiment is to roll a die with $a, b, c, d, e, f$ on each side of the die with sample space $S = \{a, b, c, d, e, f\}$

1. define a RV: $X(a) = 1, \cdots, X(f) = 6$

   $$X : S = \{a, b, c, d, e, f\} \to \overline{S} = \{1, 2, 3, 4, 5, 6\}$$

   the new random experiment is to roll a die with $1, 2, 3, 4, 5, 6$ on each side of the die
2. the old random experiment with sample space $S \Longleftrightarrow$ the new random experiment with numeric sample space $\overline{S}$
3. repeat the new random experiment is to generate a number randomly from $\overline{S} = \{1, 2, 3, 4, 5, 6\}$

# Some Conventions

- uppercase letters, e.g. $X, Y, Z \rightarrow$ RVs
- lowercase letters, e.g. $x, y, z \rightarrow$ the numeric values that RV $X, Y, Z$ can take, respectively

For a given random experiment, two probability functions are involved through $X : S \rightarrow \overline{S}$,

- $P_S(\cdot)$ is the probability function associated with $S$
- $P(\cdot)$ is the probability function associated with $\overline{S}$

$$P(X = x) \triangleq P(\{X = x\}) = P_S(\{s | X(s) = x, s \in S\})$$

$$P(X \in A) \triangleq P(\{X \in A\}) = P_S(\{s | X(s) \in A, s \in S\})$$

# Example 1, continued

The old random experiment is to roll a die with $a, b, c, d, e, f$ on each side of the die with sample space $S = \{a, b, c, d, e, f\}$

1. define a RV: $X(a) = 1, \cdots, X(f) = 6$

$$X : S = \{a, b, c, d, e, f\} \to \overline{S} = \{1, 2, 3, 4, 5, 6\}$$

   the new random experiment is to roll a die with $1, 2, 3, 4, 5, 6$ on each side of the die
2. the old random experiment with sample space $S \Longleftrightarrow$ the new random experiment with numeric sample space $\overline{S}$
3. repeat the new random experiment is to generate a number randomly from $\overline{S} = \{1, 2, 3, 4, 5, 6\}$
4. Let $x = 1$ and $A = \{1, 2\}$

$$P(X = x) \triangleq P(\{X = x\}) = P_S(\{s | X(s) = 1, s \in S\})$$

$$P(X \in A) \triangleq P(\{X \in A\}) = P_S(\{s | X(s) \in A, s \in S\})$$

# Discrete Random Variable

**Definition**

Recall that $\overline{S}$ denote the range of $X$: $\overline{S} = \{x | X(s) = x, s \in S\}$.

A RV $X$ is said to be discrete if its range $\overline{S}$ is finite or countably infinite.

# Example 1, continued

The old random experiment is to roll a die with $a, b, c, d, e, f$ on each side of the die with sample space $S = \{a, b, c, d, e, f\}$

1. define a RV: $X(a) = 1, \cdots, X(f) = 6$

$$X : S = \{a, b, c, d, e, f\} \rightarrow \overline{S} = \{1, 2, 3, 4, 5, 6\}$$

2. $X$ is discrete, because $\overline{S}$ is finite, i.e., it contains a finite number of outcomes

# Probability Mass Function (pmf)

**Definition**

Suppose that $X$ is a RV with range $\overline{S}$. Then a function $f(x) : \overline{S} \to (0,1]$ is called pmf, if

1. $f(x) > 0, \quad x \in \overline{S}$.
2. $\sum_{x \in \overline{S}} f(x) = 1$.
3. $P(X \in A) = \sum_{x \in A} f(x), \quad A \subseteq \overline{S}$,

   which defines the probability function for an event $A$. In particular, taking $A = \{x\}$ yields the probability of $X = x$, i.e.,
   $$P(X = x) = f(x)$$

# Probability Mass Function (pmf)

We often extend the domain of $f(x)$ from $\overline{S}$ to $R$ and let $f(x) = 0$, $x \notin \overline{S}$. In this case, $\overline{S}$ is called the support of $f(x)$.

**Definition**

Suppose that $X$ is a RV with range $\overline{S}$. Then a function $f(x) : R \to [0,1]$ is called pmf, if

1. $f(x) \geq 0, \quad x \in R.$
2. $\sum_{x \in \overline{S}} f(x) = 1.$
3. $P(X \in A) = \sum_{x \in A} f(x), \quad A \subseteq \overline{S}.$

# Example 1, continued

The old random experiment is to roll a die with $a, b, c, d, e, f$ on each side of the die with sample space $S = \{a, b, c, d, e, f\}$

1. define a RV: $X(a) = 1, \cdots, X(f) = 6$

$$X : S = \{a, b, c, d, e, f\} \to \overline{S} = \{1, 2, 3, 4, 5, 6\}$$

2. $X$ is discrete, because $\overline{S}$ is finite, i.e., it contains a finite number of outcomes

3. pmf $f(x) = \frac{1}{6}, \quad x \in \overline{S}$, and $f(x) = 0, \quad x \notin \overline{S}$

# Uniform Distribution

**Definition[uniform distribution]**

A RV $X$ is said to have a uniform distribution if

$$f(x) = \text{constant for} \quad x \in \overline{S}$$

# Example 2

Question: Roll a fair four-sided die twice and let $X$ be the maximum of the two outcomes. Find the pmf of $X$, $f(x)$.

# Example 2

Question: Roll a fair four-sided die twice and let $X$ be the maximum of the two outcomes. Find the pmf of $X$, $f(x)$.

1. The sample space $S$ for rolling a fair four-sided die twice is

$$S = \{(d_1, d_2) | d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}$$

2. For any $s = (d_1, d_2) \in S$, $X(s) = \max\{d_1, d_2\}$. Clearly, this RV is not one-to-one! and the range of $X$, i.e., $\overline{S} = \{1, 2, 3, 4\}$
3. To find $f(x)$, the pmf of $X$, is to find the value of $f(x) = P(X = x)$ for $x \in \overline{S}$, i.e., $x = 1, 2, 3, 4$:

$$f(1) = P(X = 1) = P_S(\{(1, 1)\}) = 1/16,$$
$$f(2) = P(X = 2) = P_S(\{(1, 2), (2, 1), (2, 2)\}) = 3/16,$$
$$f(3) = P(X = 3) = P_S(\{(1, 3), (3, 1), (2, 3), (3, 2), (3, 3)\}) = 5/16,$$
$$f(4) = P(X = 4) = P_S(\{(1, 4), (4, 1), (2, 4), (4, 2), (3, 4), (4, 3), (4, 4)\}) = 7/16,$$

# Line Graph and Probability Histogram

**Definition[Line Graph]**

A line graph of the pmf $f(x) : \overline{S} \to (0, 1]$ of a RV $X$ is a graph having a vertical line segment drawn from $(x, 0)$ to $(x, f(x))$ at each $x \in \overline{S}$

**Definition[Probability Histogram]**

If a RV $X$ with range $\overline{S}$ that only contains integers, then a probability histogram of the pmf $f(x) : \overline{S} \to (0, 1]$ is a graph having a rectangle of height $f(x)$ and a base of length 1, centered at $x$, for each $x \in \overline{S}$.

# Example 2, continued

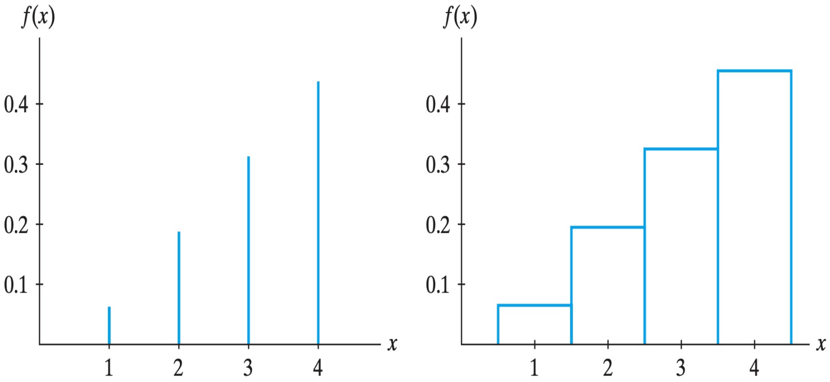



**Figure 2.1-1** Line graph and probability histogram

# Cumulative Distribution Function (cdf)

**Definition[cdf]**

The function $F(x) : R \rightarrow [0\ 1]$:

$$F(x) = P(X \leq x)$$

is called the cumulative distribution function (cdf).

1. $F(x)$ is nondecreasing and moreover,

$$P(X \leq x) = \sum_{x' \leq x, x' \in \overline{S}} f(x').$$

2. relation between the probability function and the cdf

$$P(a < X \leq b) = F(b) - F(a)$$

# Example 1, continued

The old random experiment is to roll a die with $a, b, c, d, e, f$ on each side of the die with sample space $S = \{a, b, c, d, e, f\}$

1. define a RV: $X(a) = 1, \cdots, X(f) = 6$

$$X : S = \{a, b, c, d, e, f\} \rightarrow \overline{S} = \{1, 2, 3, 4, 5, 6\}$$

2. $X$ is discrete, because $\overline{S}$ is finite, i.e., it contains a finite number of outcomes

3. pmf $f(x) = \frac{1}{6}, \quad x \in \overline{S}$, and $f(x) = 0, \quad x \notin \overline{S}$

4. cdf

$$F(x) = P(X \leq x) = \sum_{x' \leq x, x' \in \overline{S}} f(x')$$

$$= \begin{cases} 0, & x < 1 \\ \frac{k}{6}, & k \leq x < k+1, k = 1, 2, 3, 4, 5 \\ 1, & x \geq 6 \end{cases}$$

## Section 2.2 Mathematical Expectation

# Motivation

We will learn many probability distributions, it is important to introduce concepts to summarize their key characteristics.

- Mean
- Variance
- Moments
- Moment generating function

# Motivation Example

An enterprising man proposes a game: let the player throw a die and then the player receives payment as follows:

$$A = \{1, 2, 3\} \to 1 \text{ dollar}$$

$$B = \{4, 5\} \to 2 \text{ dollars}$$

$$C = \{6\} \to 3 \text{ dollars}$$

# Motivation Example

1. This defines explicitly a RV $X : S \to \overline{S}$, where $S = \{1, 2, 3, 4, 5, 6\}$ and $\overline{S} = \{1, 2, 3\}$.

$$\text{for } s \in A = \{1, 2, 3\}, \qquad X(s) = 1$$

$$\text{for } s \in B = \{4, 5\}, \qquad X(s) = 2$$

$$\text{for } s \in C = \{6\}, \qquad X(s) = 3$$

The RV $X$ represents the payment the player receives and is NOT one-to-one!

# Motivation Example, continued

2. The RV $X$ is discrete.
3. pmf of $X$:

$$f : \overline{S} \to (0,1] \quad \overline{S} = \{1,2,3\}$$

$$f(x) = \frac{4-x}{6}, \quad x = 1,2,3.$$

# Motivation Example, continued

**Question**

The man charges the player 2 dollars for each play. Can the man make profit if the game is repeated for a large number of times?

# Motivation Example, continued

4. payment of $\begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix}$ occur $\begin{Bmatrix} \frac{3}{6} \\ \\ \frac{2}{6} \\ \\ \frac{1}{6} \end{Bmatrix}$ of the times.
5. average payment is

$$1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6} = \frac{5}{3}$$

so the man can earn $2 - \frac{5}{3} = \frac{1}{3}$ per play on average

# Mathematical Expectation

More generally, we are interested in the average value of a function of $X$, say $g(X)$.

**Definition[Mathematical Expectation]**

Assume $X$ is a discrete RV with range $\overline{S}$ and $f(x)$ is its pmf. If $\sum_{x\in\overline{S}} g(x)f(x)$ exists, then it's called the mathematical expectation of $g(X)$ and is denoted by

$$E[g(X)] = \sum_{x\in\overline{S}} g(x)f(x)$$

# Motivation Example, revisited

4. payment of $\begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix}$ occur $\begin{Bmatrix} \frac{3}{6} \\ \\ \frac{2}{6} \\ \\ \frac{1}{6} \end{Bmatrix}$ of the times.

5. average payment is

$$1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6} = \frac{5}{3}$$

   so the man can earn $2 - \frac{5}{3} = \frac{1}{3}$ per play on average

6. Formally, the average payment is given by

$$E(X) = \sum_{x \in \overline{S}} x f(x) = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6} = \frac{5}{3}$$