Lecture 4: Introduction to Probability

• 4. 1 Introduction

- ***Probability*** is a numerical measure of the likelihood that an event will occur

- Probability values are always assigned on a scale from 0 to 1

    - A probability near 0 indicates an event is very unlikely to occur

    - A probability near 1 indicates an event is almost certain to occur

- ***Why Should We Care*****:** we cannot predict events with absolute certainty. Instead, we make decisions about the data we have collected using probability – our inferences are stated in probabilistic terms

    - inference is made using ***inductive*** reasoning, that is, from a limited number of observations to general rules, and the uncertainty is expressed in probabilistic terms

- *Probability* is the study of ***randomness*** and ***uncertainty***

- In the early days, probability was associated with *games of chance* (gambling)…The chance of winning …

- Some simple ***games***:

    - Game 1: A fair die is rolled. If the result is 2, 3, or 4, you win \$1; if it is 5, you win \$2; but if it is 1 or 6, you lose \$3. Should you play this game?

    - Game 2: Before two fair coins are tossed, you are given a choice of the following payoffs:
    Payoff 1: Win \$1 for each head. Lose \$3 for getting two tails.
    Payoff 2: Win \$1 if the coins are different. Win \$2 if both coins turn up tails. Lose \$3 if both coins turn up head.
    Payoff 3: Win \$3 for getting two heads. Win \$1 for getting one of each. Lose \$4 for getting two tails.
    Which payoff (if any) should you choose?

    - Game 3: Game 3: A certain lottery has a jackpot of \$10M and the chance of winning is one in 500 million. Should you play this game?

- Modern ***applications***:
    - Epidemiology
    - Weather forecast
    - Business applications: (Finance, Insurance, Marketing)

• 4. 2 Experiments, Sample Spaces, and Events

- A ***simple experiment*** is the process of obtaining observations or measurements. More formally, an experiment is some well-defined act or process that leads to a single well-defined outcome

  - Examples of simple experiments:

    - tossing a coin

    - picking a card from a deck

    - measuring temperature from patients

- The ***sample space*** for an experiment is the set of all possible experimental outcomes

- A ***sample point*, **or an ***elementary event*, **is any member of the sample space

- Any set of elementary events is an ***event***, or an ***event* class**. (e.g., H,T; spades, clubs)

  - Example: The experiment is to toss a coin 3 times

  - Sample space **S =** {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

  - Examples of event include

    A = {HHH, HHT,HTH, THH} = {at least two heads}

    B = {HTT, THT,TTH} = {exactly two tails}

• 4. 3 Probabilities

- ***Definition***: Given a sample space **S** and the family of events in **S**, a probability function associates with each event $A \subseteq \mathbf{S}$, $p(A)$, the probability of event A, such that the following axioms are true:

  - $p(A) \geq 0, A \subseteq \mathbf{S}$

  - $p(\mathbf{S}) = 1$

  - If there exists some *countable* set of events, $\{A_1, A_2, \dots, A_N\}$, and if these events are all mutually exclusive,

$$p\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} p(A_i)$$

[The probability of the union of the A_N mutually exclusive events is equal to the sum of their separate probabilities]

• 4. 4 Probability Rules

- There are some ***basic probability relationships*** that can be used to compute the probability of an event without knowledge of all the sample point probabilities

Ø	is the empty set; that is, a set that contains no events
$A \subset S$	A is a subset of S, or is fully contained in S
$A \cup B$	A union B, the set of all elements that are either in A or B or both ($A \cup B = S$)
$A \cap B$	A intersection B, the set of all elements that are in both A and B (if A and B are mutually exclusive, $A \cap B =$ Ø)
A^c	A complement; that is, the event consisting of all sample points that are not in A (if A and B are mutually exclusive, $A^c = B$ and $B = A^c$, $\sim A$, or $\overline{A}$)
$A \cup A^c = S$	A union its complement is S

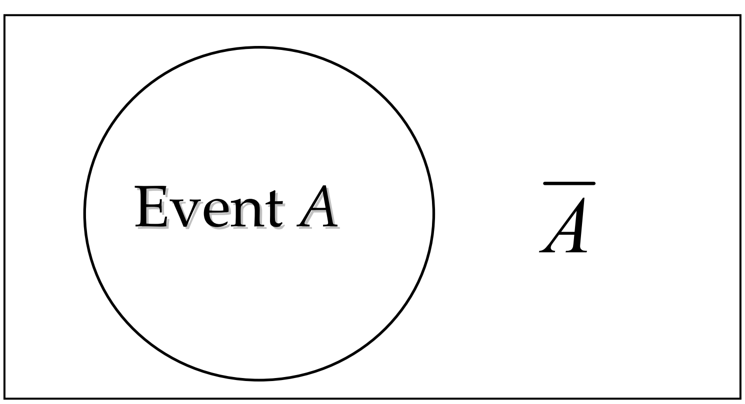
- ***Rule 1***: *The complement rule of probability*

- The ***complement*** of an event A is defined to be the event consisting of all sample points that are not in A

- $p(\sim A) = 1 - p(A)$

- The complement of A is denoted by $\sim A$ or A^c or $\overline{A}$

- The ***Venn diagram*** below illustrates the concept of a complement



- ***Rule 2***: *Probability Range*

$0 \leq p(A) \leq 1$

- ***Rule 3***: *Rule of the impossible event*

$p(\varnothing) = 0$, for any $\boldsymbol{S}$

- ***Rule 4***: *The "or" rule of probability*

- The probability that event A or B (both in ***S***) will occur is
$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

where $P(A \cap B)$ is the probability that both events, *A* and *B,* will occur

- *A* and *B* are ***mutually exclusive events*** if both cannot occur together in the same experiment
$p(A \cap B)=0$

- The ***Additional Rule for Mutually Exclusive Events***:

If *A, B, C,* … are mutually exclusive events, then

$p(A \cup B \cup C \cup \ldots) = p(A)+p(B)+p(C)+ \ldots$

- ***The Multiplication Rule of Probability***

- The probability that both *A* and *B* will occur when an experiment is performed is given by

$p(A \cap B) = p(A) \cdot p(B|A)$,

where $p(B|A)$ is the probability of B if A has occurred, and is called the **conditional probability** of B given A

- Events *A* and *B* are ***independent*** if the probability of each event is not affected by whether or not the other event has occurred

- ***The Multiplication Rule for Independent Events*:**

- If *A, B, C,* … are independent events, then

$p(ABC \ldots) = p(A) \cdot p(B) \cdot p(C) \cdot \ldots$

- ***Test for Independent Events***:

- A and B are ***independent*** if either $p(\mathrm{A}|\mathrm{B}) = p(\mathrm{A})$ or $p(\mathrm{AB})=p(\mathrm{A}) \bullet p(\mathrm{B})$. Otherwise, the events are dependent

- ***The Condition Rule of Probability***

- The conditional probability of B, given that A has occurred, is given by

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

Provided that $A \neq 0$

• 4. 5 Assigning Probabilities

- ***Relative Frequency Method*** - Assigning probabilities based on experimentation or historical data

  - The probability of event E is denoted by $P(E)$ and is the proportion of the time that E can be expected to occur ***in the long run***

  - If we let A be the event

    A: a female birth in the United States

    then the probability of E is estimated to be

$$P(E) \approx \frac{1{,}983{,}000}{4{,}065{,}000} = .49$$

- ***Bernoulli's Theorem***
- If the probability of occurrence of the event X is $p(X)$ and if N trials are made, independently and under exactly the same conditions, the probability that the relative frequency of occurrence of X differs from $p(X)$ by any amount, however small, approaches zero as the number of trial grows indefinitely large

$$\mathrm{Lim}_{N \to \infty} F(X) - p(X) = 0,$$

where $F(X)$ is the relative frequency of X to N

- ***Classical/Theoretical Method*** - Assigning probabilities based on the assumption of equally likely outcomes

- Suppose a sample space has ***S*** equally likely points, and *A* is an event consisting of *a* points. Then

$$p(A) = \frac{a}{s}$$

- The sample space of the experiment is

(a,a)	(a,b)	(a,c)
(b,a)	(b,b)	(b,c)
(c,a)	(c,b)	(c,c)

The sample space has s = 9 equally likely outcomes

- The event of interest is
A: both answers are correct

Event *A* contains a=1 point. Therefore, by the classical method of assigning probabilities, the probability that the student will answer both questions correctly is

$$P(A) = \frac{a}{s} = \frac{1}{9}$$

- ***Subjective Method*** - Assigning probabilities based on the assignor's judgment

  - When economic conditions and a company's circumstances change rapidly it might be inappropriate to assign probabilities based solely on historical data

  - We can use any data available as well as our experience and intuition, but ultimately a probability value should express our ***degree of belief*** that the experimental outcome will occur

  - The best probability estimates often are obtained by combining the estimates from the classical or relative frequency approach with the subjective estimates

• 4. 6 Probability Distributions

- A ***probability distribution*** is any statement of a function associating each of a set of mutually exclusive and exhaustive events with its probability

- The sum of the probabilities of all classes must be 1.00

Probability Distribution

Height (inches)	p
78-82	.002
73-77	.021
68-72	.136
63-67	.682
58-62	.136
53-57	.021
48-52	.002
	1.000

- Given a frequency distribution, each interval has associated a probability, as

$p(A_I) = f(A_I) / f(A_N)$,

or the frequency of a particular interval divided by the total frequency

- Given a theoretical probability distribution and N observations made independently and at random with replacement, there is a ***theoretical frequency distribution***, in which for any event class A

$F(A)$ from $N = Np(A)$,

where $F(A)$ is the theoretical frequency of event A out of N observations

Theoretical Frequency Distribution

Height (inches)	*f*
78-82	2
73-77	21
68-72	136
63-67	682
58-62	136
53-57	21
48-52	2
	1,000 = N

- **4. 7 Random Variables**

- Let X be a function that associates a real number with each and every elementary event in some sample space $\boldsymbol{S}$. Then X is called a ***random variable*** on the sample space $\boldsymbol{S}$

- A random variable X represents values that are associated with elementary events, so that particular values of X occur when the appropriate elementary events occur, however the association might be

- Random variables can be specified in three ways: listing all possible numerical events and their associated probability, graphing this relationship, or expressing a rule of the probability for each value

- Typical notation is to use capital letters, such as X, Y, Z to denote random variables and lowercase letters, such as x, y, z to denote particular values of the random variable

- A variable X is said to be a ***discrete random variable*** if it can assume only a particular finite or a countable infinite set of values. Given discrete random variables, probability calculations are often simple

- Example: two dices, and $X = \text{die}_1 + \text{die}_2$

*Values of X for all elementary event in **S***

6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	**2**	**3**	**4**	**5**	**6**

- From the definition of X we can now estimate the probability distribution of all values x of X, as

Probability Distribution for X

X	p
12	1/36
11	2/36
10	3/36
9	4/36
8	5/36
7	6/36
6	5/36
5	4/36
4	3/36
3	2/36
2	1/36
	36/36

- And from this distribution of X, one can compute other probability questions

- Example 1: $p(3 \leq X \leq 5)$?

$p(3 \leq X \leq 5) = p(3 \text{ U } 4 \text{ U } 5) = p(3) + p(4) + p(5)$

$= p(X = 3) + p(X = 4) + p(X = 5)$

$= 2/36 + 3/36 + 4/36 = 9/36 = 1/4 = .25$

- Example 2: $p(X < 5)$?

$p(X < 5) = p(X = 2) + p(X = 3) + p(X = 4)$

$= 1/36 + 2/36 + 3/36 = 6/36 = 1/6$

• 4. 8 Function Rules for Discrete Random Variables

- Sometimes the most suitable way to specify the distribution of a random variable is by its rule or function

- Example 1: Let X be a random variable that can take one of six values 1, 2, … , 6. If all values have exactly the same probability of occurrence, the function rule for X can be expressed as

$$p(x) = \begin{cases} 1/6 & (\text{if } x = 1, 2, \ldots, 6) \\ 0 & (\text{otherwise}) \end{cases}$$

- Example 2: Let X be a random variable that can take one of six values 1, 2, … , 6. Let the function rule for X be expressed as

$$p(x) = \begin{cases} x/12 & (\text{if } x = 1, 2, 3) \\ (7-x)12 & (\text{if } x = 4,5,6) \\ 0 & (\text{otherwise}) \end{cases}$$

What are the probabilities to values of X?

$p(X = 1) = 1/12$; $p(X = 2) = 2/12$; $p(X = 3) = 3/12$;

$p(X = 4) = (7\text{-}4)/12 = 3/12$;

$p(X = 5) = (7\text{-}5)/12 = 2/12$;

$p(X = 6) = (7\text{-}6)/12 = 1/12$;

• 4. 9 Continuous Random Variables

- A random variable is ***continuous*** when it can be represented in terms of arbitrarily small class intervals of size ΔX, with the probability of any interval corresponding exactly to the area cut off by the interval under a smooth curve

- As ΔX approaches 0, the p associated with any class interval also approaches 0, because the corresponding area under the curve is being reduced

- Hence, the occurrence of ***any exact value*** of X is said to have zero probability

- And probabilities are not discussed for a particular value of a continuous random variable X, but for intervals of X in a continuous distribution

- Similarly, for a particular value a, probability is described as the ***probability density*** of X at value a, expressed as

$f(a)$ = probability density of X at a

- The probability density of X at a is the ***rate of change*** in the p of an interval with lower limit a, for very small changes in the size of the interval

- This rate of change depends on:

  - the function rule assigning probabilities to intervals

  - the particular region of X values of interest

- For discrete random variables, probability at value x is equivalent to density at value x; this is not true for continuous random variables

- The probability of an interval for a continuous random variable depends on the weighted sum of the probability densities associated with all the values in the interval, as

$$p(a \leq X \leq b) = \int_a^b f(x)dx.$$

or the area cut off by the interval under the curve for the probability densities, when X is a continuous random variable and the total area = 1

- The distribution of a random variable can also be described using the ***cumulative distribution function***. This function represents the relation between the possible values a of a random variable X and the probability that the value of X is less than or equal to a

  - The cumulative probability $F(a) = p(X \leq a)$

  - Similarly, $F(b) - F(a) = p(a \leq X \leq b)$

- Graphic representation of continuous distributions (probability densities, areas, and probabilities in a continuous distribution).