Stochastic Signal Processing

Lesson 1: Basic of Probability and Random Variables

Weize Sun

PROBABILITY THEORY - Basics

- Probability theory deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them.

- The notion of an experiment assumes a set of repeatable conditions that allow any number of identical repetitions.

- When an experiment is performed under these conditions, certain elementary events occur in different but *completely uncertain* ways.

PROBABILITY THEORY - Basics

- The probability of an event（事件） A is defined as:

$$P(A) \cong n_A/n$$ *(≅ means approximate with a high degree of certainty)*

where n_A is the number of occurrences of A and n is the total number of trials. (When $n \to \infty$, it is equal)

- The totality(总体) of all ξ_i, *known a priori*, constitutes a set S, the set of all experimental outcomes.

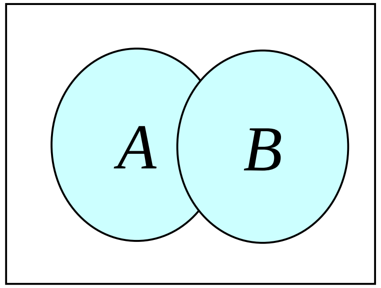
$$S = \{\xi_1, \xi_2, \dots \xi_k, \dots\}$$

PROBABILITY THEORY - Basics

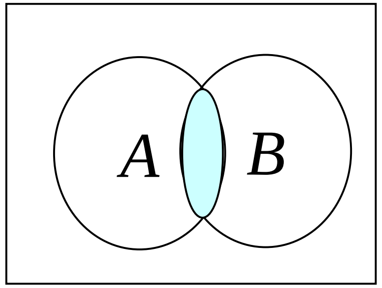
- S has subsets A, B, C. If A is a subset of S, then $\xi \in A$ implies $\xi \in S$.
- And:

$$A \cup B = \{\xi \mid \xi \in A \;\; \text{or} \;\; \xi \in B\}$$
$$A \cap B = \{\xi \mid \xi \in A \;\; \text{and} \;\; \xi \in B\}$$

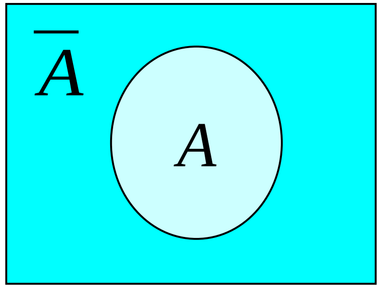
$$\overline{A} = \{\xi \mid \xi \notin A\}$$



$A \cup B$



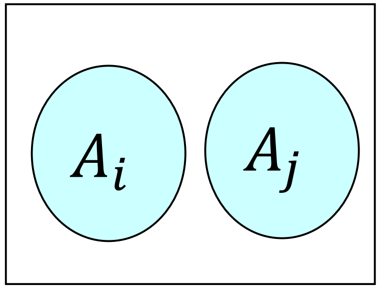
$A \cap B$



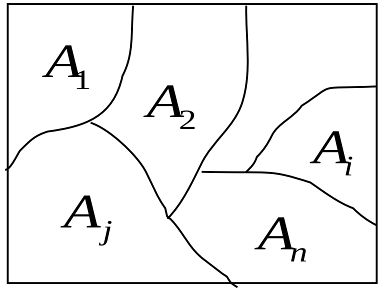
$\overline{A}$

PROBABILITY THEORY - Basics

- If $A \cap B = \emptyset$, the empty set, then A and B are said to be mutually exclusive (M.E).
- A partition(分割) of S is a collection of mutually exclusive subsets $\{A_k\}$ of S such that their union(并/并集) is S.



$$A_i \bigcap A_j = \emptyset \; for \; i \neq j$$



$$\bigcup_i A_i = S$$

PROBABILITY THEORY - Basics

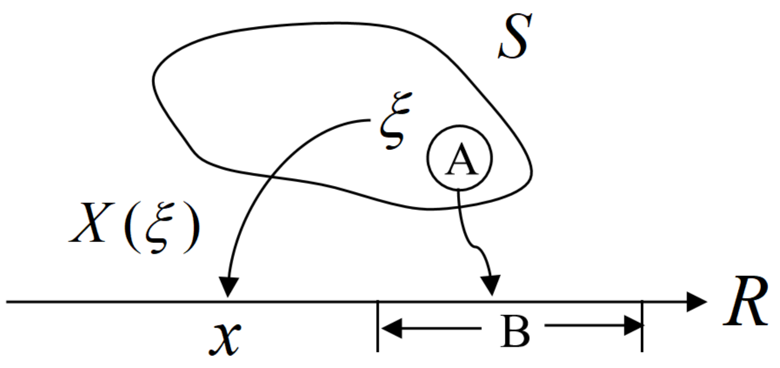
- Suppose two subsets *A* and *B* are both events, then consider
  - "Does an outcome belongs to *A* or *B?"*
  - "Does an outcome belongs to *A* and *B?"*
  - "Does an outcome falls outside *A"*?
- Thus the sets $A \cup B,\ A \cap B,\ \overline{A},\ \overline{B},$ etc., also qualify as events.

Before we go on to probability, let's get a quick understanding of

"what is **Stochastic**"

Random Variables - Basics

- Let S *is the set*, *and* X is a function that maps every even $\xi \in S$, to a unique point $x \in R,$ the set of real numbers.
- Since the outcome ξ is not certain, so is the value $x \in R$.
- If B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$".
- To determine this probability, we can look at the set $A = X^{-1}(B) \in S$ that contains all $\xi \in S$ that maps into B under the function X.



Random Variables - Basics

- Then: Probability of the event $"X(\xi) \in B" = P(X^{-1}(B))$.
- The notion(概念) of random variable (r.v) makes sure that the inverse mapping $X^{-1}(B)$ always results in an event so that we are able to determine the probability for any $B \in$ R:

- **Random Variable (r.v)**: A finite single valued function that maps the set of all experimental outcomes S into the set of real numbers R is said to be a r.v, if the set $\{\xi | X(\xi) \leq x\}$ is an event for every *x* in *R*.
- Note: usually we right $\{\xi | X(\xi) \leq x\}$ as $\{X \leq x\}$
- This is what we discussed in ‘Statistic（统计学）’

What is Stochastic?

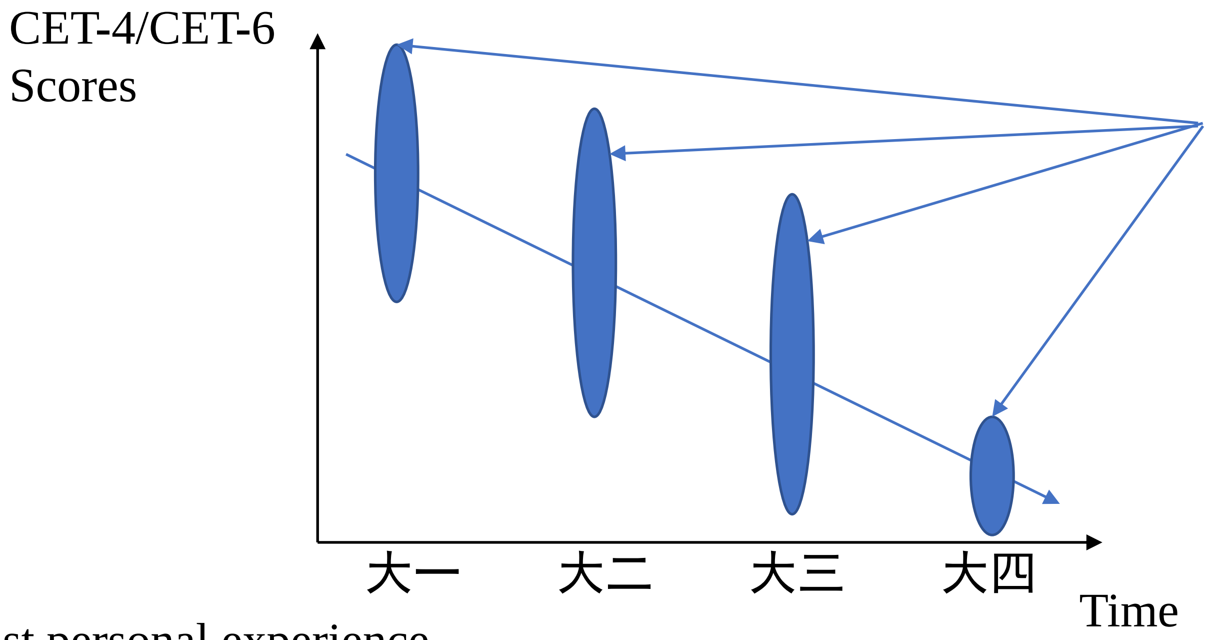
– A word from the very beginning

- Probability and random variable deal with things that are random, but its randomness will not change, for example:
  - Flipping a coin
  - Rolling a dice
  - Me taking a driver’s license exam – will always fails, 100% sure, therefore it is a deterministic event（确定性事件）
  - …

What is Stochastic?

– A word from the very beginning

- **Stochastic** deals with things that change according to time and state:
  - Your CET-4/CET-6 Scores:
  - Statistic told us that a person can get his/her highest CET-4/CET-6 Scores right after the college entrance examination（高考）*



The probability of a person’s CET-4/CET-6 Scores. For example, in Year 1, he might get scores 710 – 510, but in Year 4, he might get only 450-300.

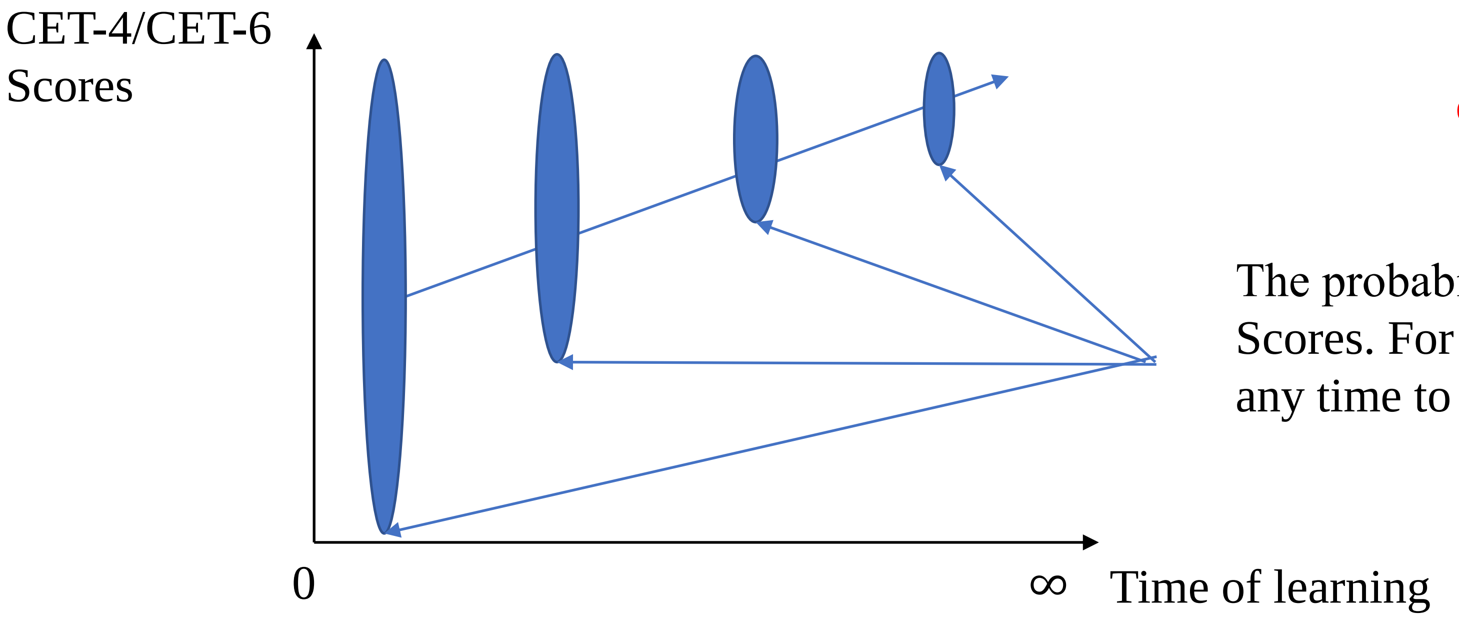
change according to time

*: just personal experience

What is Stochastic?

– A word from the very beginning

- **Stochastic** deals with things that change according to time and state:
  - Your CET-4/CET-6 Scores:
  - Statistic told us that a person can get higher CET-4/CET-6 Scores under hard learning (comparing to himself only)



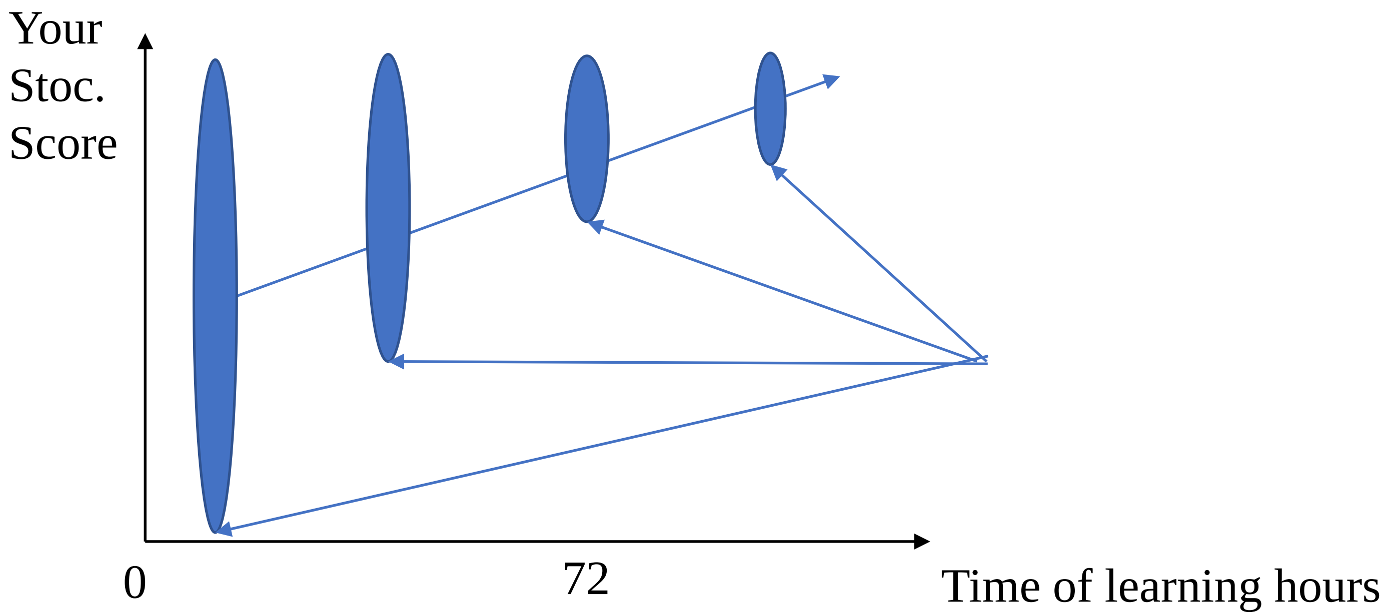
change according to state

The probability of a person’s CET-4/CET-6 Scores. For example, if a person do not pay any time to learn it, he might get 710-300.

What is Stochastic?

– A word from the very beginning

- Listen carefully, finish homework and experiments by yourself
- And will get a high score!

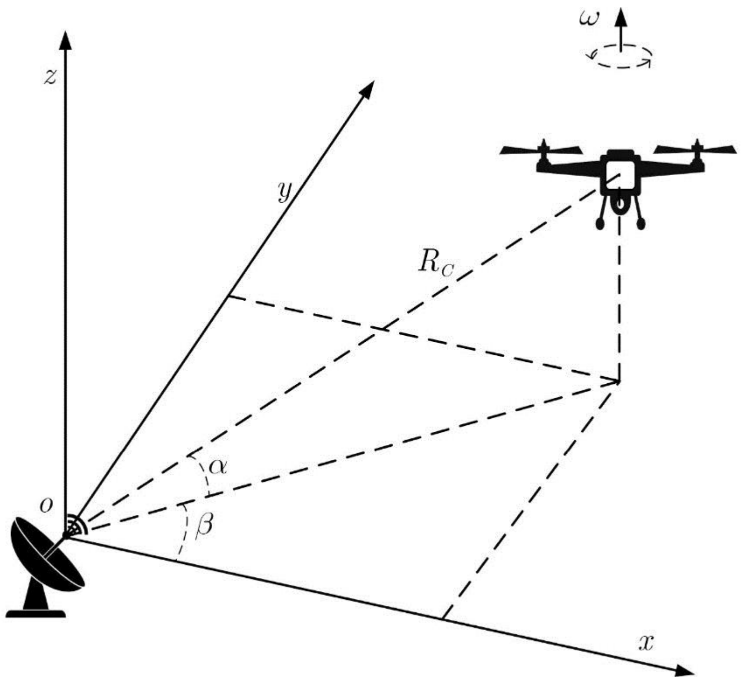


概率分析概率分，
混沌理论混沌里；
随机过程随机过，
量子力学量力学；
实变函数学十遍，
泛函分析心犯寒；
微机原理闹危机，
汇编语言不会编；
……

What is Stochastic?

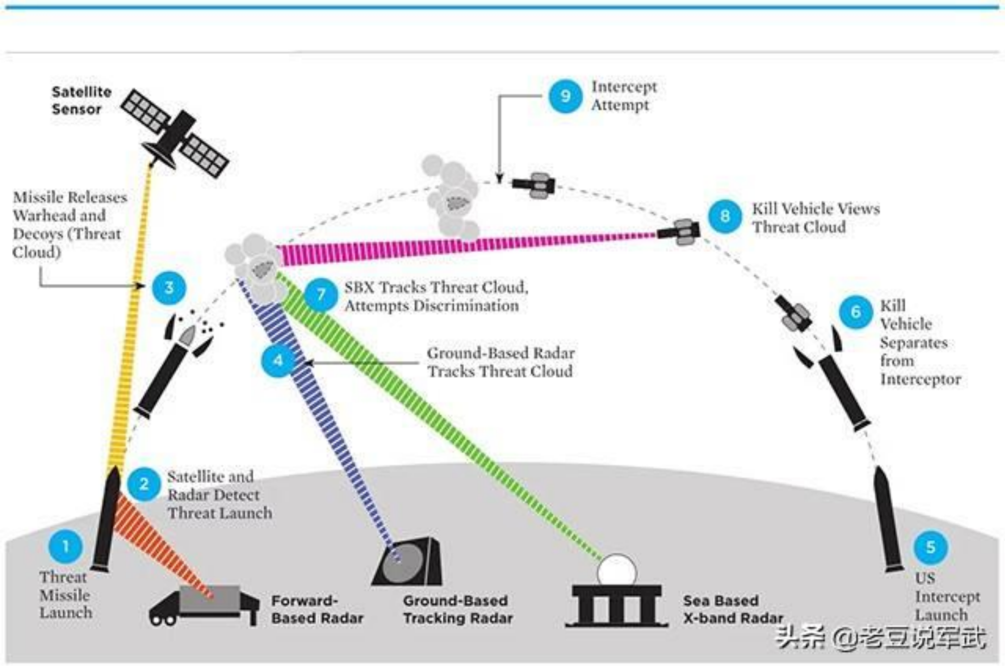
– A word from the very beginning

- **Stochastic** deals with things that change according to time and state
- **Stochastic** is the tool to analyze such ‘time or state varying probability’
    - Correlation(相关性) and Power spectrum(功率谱): mainly focus on time varying



Radar:

- sent out signal
- Receive the echo signal （回波信号）
- Use Time-Frequency analysis to find where the objects are
- Will try a small case of this in later experiment (experimental report required)



What is Stochastic?

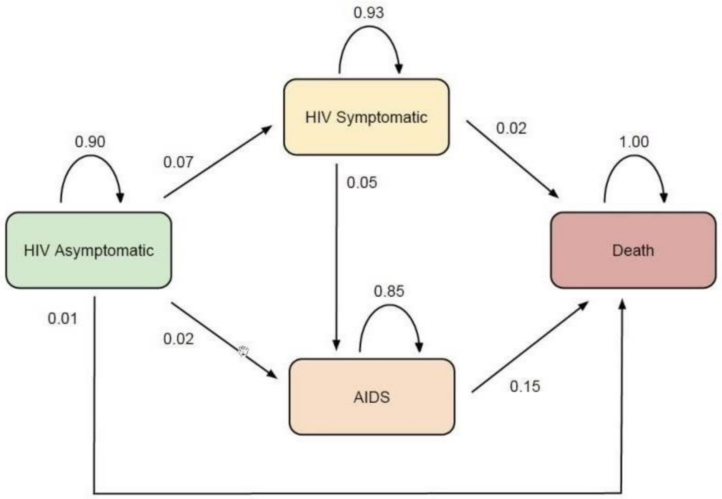
– A word from the very beginning

- **Stochastic** deals with things that change according to time and state
- **Stochastic** is the tool to analyze such 'time or state varying probability'
  - Markov: mainly focus on state varying
  - Example: HIV expectation
  - HIV has these states:
    - HIV asymptomatic（无临床症状）
    - HIV symptomatic（有临床病状）
    - AIDS（获得性免疫缺陷综合征）
    - Death

This stochastic matrix:

$$\begin{pmatrix} 0.90 & 0.07 & 0.02 & 0.01 \\ 0 & 0.93 & 0.05 & 0.02 \\ 0 & 0 & 0.85 & 0.15 \\ 0 & 0 & 0 & 1.00 \end{pmatrix}$$

corresponds to:



Now we go back to probability

Axioms(公理) of Probability

- For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the axioms of probability.
  I. $P(A) \geq 0$ (Probability is a nonnegative number)
  II. $P(S) = 1$ (Probability of the whole set is unity)
  III. If $A \cap B = \phi$, then $P(A\ U\ B) = P(A) + P(B)$ (1-1)

(Note that (III) states that if A and B are mutually exclusive (互不相容，M.E.) events, the probability of their union is the sum of their probabilities.)

Conditional Probability and Independence

- In N independent trials, suppose N_A, N_B, N_{AB} denote the number of times events A, B and AB occur respectively. According to the frequency interpretation of probability, for large N

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}.$$

- Among the N_A occurrences of A, only N_{AB} of them are also found among the N_B occurrences of B. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$$

is a measure of "the event A given that B has already occurred".

Conditional Probability and Independence

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$$

- is a measure of "the event A given that B has already occurred". We denote this conditional probability by

  $P(A|B) =$ Probability of "the event A given that B has occurred".

- We define

$$P(A|B) = \frac{P(AB)}{P(B)}, \qquad (1\text{-}2)$$

- provided $P(B) \neq 0$. Note that the above definition satisfies all probability axioms discussed earlier.

We have

(i) $P(A|B) = \dfrac{P(AB) \geq 0}{P(B) > 0} \geq 0,$

III. If $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$ (1-1)

$$P(A|B) = \frac{P(AB)}{P(B)}, \qquad (1\text{-}2)$$

(ii) $P(S|B) = \dfrac{P(SB)}{P(B)} = 1$, since $S \cap B = B$.

(iii) Suppose $A \cap C = \phi$, Then

$$P(A \cup C \mid B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}.$$

But $AB \cap CB = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB)$.

$$P(A \cup C \mid B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A|B) + P(C \mid B),$$

Properties of Conditional Probability:

a. If $B \subset A, \quad AB = B,$ and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

since if $B \subset A,$ then occurrence of B implies automatic occurrence of the event A. As an example

$$A = \{\text{outcome is even}\}, \quad B = \{\text{outcome is 2}\},$$

in a dice tossing experiment. Then $B \subset A,$ and $P(A \mid B) = 1.$

Also, if $A \subset B, \quad AB = A,$

Properties of Conditional Probability:

b. If $A \subset B, \quad AB = A,$ and

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

(In a dice experiment, $A =$ {outcome is 2}, B={outcome is even}, so that $A \subset B.$ The statement that B has occurred (outcome is even) makes the odds for "outcome is 2" greater than without that information).

Properties of Conditional Probability:

c. We can use the conditional probability to express the probability of a complicated event in terms of "simpler" related events.

Let $A_1, A_2, \cdots, A_n$ are pair wise disjoint and their union is S. Thus $A_i A_j = \phi$, and

$$\bigcup_{i=1}^{n} A_i = S$$

Thus $B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n .$

Properties of Conditional Probability:

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n .$$

- But $A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi,$ so that:

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i). \quad （全概率公式）$$

- With the notion of conditional probability, next we introduce the notion of “independence” of events.

- **Independence**: *A* and *B* are said to be independent events, if

$$P(AB) = P(A) \cdot P(B).$$

- Notice that the above definition is a probabilistic statement, *not* a set theoretic notion （集合论的概念） such as mutually exclusiveness.

Properties of Conditional Probability:

- Suppose A and B are independent, then

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

- Thus if A and B are independent, the event that B has occurred cannot provide extra information whether the event A occurred, or say, it makes no difference to A whether B has occurred or not. An example will clarify the situation:

Properties of Conditional Probability:

Example 1: A box contains 6 white and 4 black balls. Remove two balls at random without replacement. What is the probability that the first one is white and the second one is black?

Let W_1 = "first ball removed is white"

B_2 = "second ball removed is black"

Properties of Conditional Probability:

Example 1: A box contains 6 white and 4 black balls. Remove two balls at random without replacement. What is the probability that the first one is white and the second one is black?

Let W_1 = "first ball removed is white"

B_2 = "second ball removed is black"

- We need $P(W_1 \cap B_2) = ?$ We have $W_1 \cap B_2 = W_1B_2 = B_2W_1$. Using the conditional probability rule,

$$P(W_1B_2) = P(B_2W_1) = P(B_2 \mid W_1)P(W_1).$$

$$P(W_1) = \frac{6}{6+4} = \frac{6}{10} = \frac{3}{5}, \quad P(B_2 \mid W_1) = \frac{4}{5+4} = \frac{4}{9}, \Rightarrow P(W_1B_2) = \frac{3}{5} \cdot \frac{4}{9} = \frac{12}{45}$$

Properties of Conditional Probability: $P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i).$

Are the events W_1 and B_2 independent?

Our common sense says No.

Properties of Conditional Probability: $P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i).$

Are the events W_1 and B_2 independent?

To verify this we need to compute $P(B_2)$. Of course the fate of the second ball very much depends on that of the first ball. The first ball has two options: W_1 = "first ball is white" or B_1= "first ball is black". Note that $W_1 \cap B_1 = \emptyset$ and $W_1 \cup B_1 = S$. Hence W_1 together with B_1 form a partition. Thus:

$$P(B_2) = P(B_2 \mid W_1)P(W_1) + P(B_2 \mid R_1)P(B_1)$$

$$= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{4+2}{15} = \frac{2}{5},$$

$\Rightarrow$ $P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(W_1 B_2) = \frac{12}{45}$ $\Rightarrow$ events W_1 and B_2 are not dependent

Bayes' theorem

PPT page 18:
$$P(A|B)=\frac{P(AB)}{P(B)},$$

$$P(AB)=P(A|B)P(B).$$

$$P(B|A)=\frac{P(BA)}{P(A)}=\frac{P(AB)}{P(A)},$$

$$\Rightarrow \quad P(AB)=P(B|A)P(A).$$

or
$$P(A|B)P(B)=P(B|A)P(A).$$

known as Bayes' theorem:

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

provided $P(B)\neq 0$. As we show below, the above definition satisfies all probability axioms discussed earlier.

Bayes' theorem

$$P(A\,|\,B) = \frac{P(B\,|\,A)}{P(B)} \cdot P(A)$$

- Although simple enough, Bayes' theorem has an interesting interpretation:
  - $P(A)$ represents the a-priori（先验） probability of the event A. Suppose B has occurred, and assume that A and B are not independent. How can this new information be used to update our knowledge about A?
  - Bayes' rule will take into account the new information ("B has occurred") and gives out the a-posteriori （后验） probability of A given B.
  - We can also view the event B as new knowledge obtained from a fresh experiment. We know something about A as $P(A)$. The new information is available in terms of B.
  - The new information should be used to improve our knowledge/understanding of A. Bayes' theorem gives the exact mechanism for incorporating such new information.

Bayes' theorem

A more general version of Bayes' theorem involves partition of S:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)},$$

$A_i,\ i = 1 \to n,$ represent a set of mutually exclusive events with associated a-priori probabilities $P(A_i),\ i = 1 \to n.$ With the new information "B has occurred", the information about A_i can be updated by the n conditional probabilities

Bayes' theorem

- Example 2: Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) has 15 defective bulbs and the second 5. Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

(b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1?

Try it yourself

Bayes' theorem

- Example 2: Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) has 15 defective bulbs and the second 5. Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

- Solution: Note that box B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs. Let D = "Defective bulb is picked out":

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025.$$

Bayes' theorem

- Since a box is selected at random, they are equally likely:

$$P(B_1) = P(B_2) = \frac{1}{2}.$$

- Thus B_1 and B_2 form a partition, and

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$

$$= 0.15 \times \frac{1}{2} + 0.025 \times \frac{1}{2} = 0.0875 .$$

- Thus, there is 8.75% probability that a bulb picked at random is defective.

Bayes' theorem

- (b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1? $P(B_1 \mid D) = ?$

$$P(B_1 \mid D) = \frac{P(D \mid B_1)P(B_1)}{P(D)} = \frac{0.15 \times 1/2}{0.0875} = 0.8571 \; .$$

- Notice that initially $P(B_1) = 0.5;$ then we picked out a box at random and tested a bulb that turned out to be defective. Can this information shed some light about the fact that we might have picked up box 1?

- We can see that $P(B_1 \mid D) = 0.857 > 0.5,$ and indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall box1 has six times more percentage of defective bulbs compared to box2).