Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Sep. 23 2021

Ve501 2021-2022 Fall

Based on Lecture Notes by Prof. Yong Long

Outline

- 1. Introduction to Probability
  - Application example
  - Review of set and functions
  - Models of random experiments
  - Axioms and properties of probability
  - Conditional probability
  - Independence of events
  - Combinatorics and probability

Conditional probability

Conditional Probability: Definition

- The conditional probability $\mathrm{P}(A|B)$ of event A occurring given that event B has occurred is a model for the conditional frequency with which A occurs when B has also occurred.

- Example: dice toss
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = 2^{\Omega}$, $\mathrm{P}(\mathrm{A}) = \frac{|\mathrm{A}|}{6}$, X
  - $A=\{3\}$, $B = \{1, 3, 5\}$
  - Find $\mathrm{P}(A|B)$

Example

x1	x2	x3	x4	x5	x6	x7	x8	x9	X10
3	2	6	4	2	5	4	1	1	3
A									A
B					B		B	B	B

- Frequency of A is 2/10
- Frequency of B is 5/10
- Frequency of A given B is 2/5

- $P(A|B) = 2/5$?

Conditional Probability: Relationship

- Relationship to "unconditional" probability

$$P(A|B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)}, P(B) \neq 0 \\ \text{undefined}, P(B) = 0 \end{cases}$$

- usually apply

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

  – $P(A \cap B)$ = **joint probability** of A and B
  – $P(A|B)$ = conditional probability of A given B

- Previous examples:
  - $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/6}{3/6} = \dfrac{1}{3}$

Example

- There are 200 light bulbs in one box. Each light bulb is good or defective. Three are three different powers: 50w(watt), 100w, or 150w

	50W	100W	150W
G	85	63	27
D	15	7	3

- Experiment: draw a bulb at random from the box

$$\Pr(D|50W) = ? \qquad \Pr(50W|D) = ?$$

Solution

$$P_r(D|50w) = \frac{P_r(\textit{defective of } 50w)}{P_r(50w)} = \frac{15/200}{100/200} = \frac{15}{100} = \frac{3}{20}$$

$$P_r(50w|D) = \frac{P_r(\textit{defective of } 50w)}{P_r(\textit{defective})} = \frac{15/200}{25/200} = \frac{15}{25} = \frac{3}{5}$$

Conditional Probability: Properties

- Conditional probability behaves like ordinary probability
  - It satisfies all the axioms and all the consequences of all axioms

$$\mathrm{P}(A \cup B|C) = \mathrm{P}(A|C) + \mathrm{P}(B|C)$$

  - if A and B are disjoint
  - $\mathrm{P}(A|B)$ is undefined if $(B) = 0$
- Question: which of the following is true
  - $\mathrm{P}(A|B) \geq \mathrm{P}(A)$
  - $\mathrm{P}(A|B) \leq \mathrm{P}(A)$
  - **Both wrong**

The Law of Total Probability

- The law of total probability gives an expression for computing the probability of an event that can occur in different ways

- Let $\{B_1, B_2, \ldots, B_n\}$ be a partition of Ω, and let $B_1, B_2, \ldots, B_n \in \mathcal{A}, \mathrm{P}(B_k) > 0, k = 1, 2, \ldots, n$, then we can write

$$\mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A \cap B_k) = \sum_{k=1}^{n} \mathrm{P}(A|B_k)\,\mathrm{P}(B_k)$$

- This expression is the law of total probability.

Proof

$$A = A \cap \Omega = A \cap \left(\bigcup_{k=1}^{n} B_k \right)$$

$$= \bigcup_{k=1}^{n} (A \cap B_k) \quad (generalizeddistributivelaw)$$

Observe $(A \cap B_i)$ *and* $(A \cap B_j)$ *are pairwise disjoint for* $i \neq j$

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$$

($\{B_1, B_2, \ldots, B_n\}$ *is a partition of* Ω*)*

$$\mathrm{P}(A) = \mathrm{P}\left(\bigcup_{k=1}^{n} (A \cap B_k) \right) = \sum_{k=1}^{n} \mathrm{P}(A \cap B_k) \quad (Kolmogorov\text{'s axioms } 3)$$

Example

- 80% of students who do HW pass the exam
- 10% of students who do not do HW pass the exam
- 60% of students who do HW
- suppose the student is randomly chosen. What is the probability that the student pass the exam?

$$\mathrm{P_r}(HW) = 0.6, \quad \mathrm{P_r}(P|HW) = 0.8, \quad \mathrm{P_r}(P|no\ \ HW) = 0.1$$

$$\begin{aligned}
\mathrm{P_r}(P) &= \mathrm{P_r}((P \cap HW)\ or\ (P \cap no\ \ HW)) \\
&= \mathrm{P_r}(P \cap HW) + \mathrm{P_r}(P \cap no\ \ HW) \\
&= \mathrm{P_r}(HW)\,\mathrm{P_r}(P|HW) + \mathrm{P_r}(no\ \ HW)\,\mathrm{P_r}(P|no\ \ HW) \\
&= 0.6 \cdot 0.8 + (1 - 0.6) \cdot 0.1 = 0.52
\end{aligned}$$

Bayes' Rule

- Bayes' rule provides an expression for the conditional probability of an event A given B

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)\,P(A^c)}$$

Proof. *Combine the expression for the conditional probability*

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

with the law of total probability

$$P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$$

Generalizing Law of Total Probability

- Let $A_1, A_2, \dots, A_n$ be a set of **pairwise disjoint** events, $A_k \in \mathcal{A}, k = 1, 2, \dots, n$, such that

$$\sum_{k=1}^{n} \mathrm{P}(A_k) = 1$$

- Then for any event B

$$\mathrm{P}(B) = \sum_{k=1}^{n} \mathrm{P}(B|A_k)\,\mathrm{P}(A_k)$$

- Do $A_1, A_2, \ldots, A_n$ form a partition of Ω?
  - Since we do not require $\bigcup_{k=1}^{n} A_k = \Omega$, then A_k do not, strictly speaking, for a partition. The reminder set $A^c = (\bigcup_{k=1}^{n} A_k)^c$ has probability zero

General Form of Bayes' Rule

- Applying the general form of the law of total probability yields the general form of Bayes' rule

$$\mathrm{P}(A_k|B) = \frac{\mathrm{P}(B|A_k)\mathrm{P}(A_k)}{\sum_{k=1}^{n} \mathrm{P}(B|A_k)\mathrm{P}(A_k)}$$

  - The probabilities $\mathrm{P}(A_k), k = 1, 2, \dots, n$ are called prior probabilities of events A_k, $k = 1, 2, \dots, n$
  - The probabilities $\mathrm{P}(A_k|B)$, $k = 1, 2, \dots, n$ are called posterior probabilities of events A_k, $k = 1, 2, \dots, n$ given B

General Form of Bayes' Rule

- It is useful when there is a cause & effect relationship.
  - Some event happens that we do not observe, but for which we have some statistical information- prior probabilities $\mathrm{P}(A_k)$
  - Based on the event B we observed and the conditional probability $\mathrm{P}(B|A_k)$ we know , we would like to find the posterior probabilities $\mathrm{P}(A_k|B)$ of event A_k conditioned on the observation of B

Example

- Blood test for disease X
  - test is 98% accurate for people who have X
  - test is 97% accurate for people who do not have X
  - P = test is positive, NP = test is negative
  - 1% of people have disease X
- Your test result is P, what's the probability you have disease X?

Solution

- D=have disease X, ND= do no have disease X
- Solve for $\mathrm{P}(D|P)$
- We have $\mathrm{P}(D) = 0.01, \mathrm{P}(P|D) = 0.98, \mathrm{P}(NP|ND) = 0.97$

$$\mathrm{P}(D|P) = \frac{\mathrm{P}(P|D)\mathrm{P}(D)}{\mathrm{P}(P)} = \frac{\mathrm{P}(P|D)\mathrm{P}(D)}{\mathrm{P}(P|D)\mathrm{P}(D) + \mathrm{P}(P|ND)\mathrm{P}(ND)}$$

$$= \frac{0.98 * 0.01}{0.98 * 0.01 + (1 - 0.97)0.99} = 0.248$$

Chain Rule

$$P(A_1 \cap A_2 \ldots \cap A_n)$$
$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \ \cap \ A_2) \ldots P(A_n|A_1 \cap A_2 \ldots \cap \ A_{n-1})$$

$P(A_1 \cap A_2 \ldots \cap A_n)$=$P(A_1)P(B|A_1)$, where $B \triangleq A_2 \cap \cdots \cap A_n$

Example

- In a factory there are 100 units of a certain product, 5 of which are defective. We pick three units from the 100 units at random. What is the probability that none of them are defective?

- Solution:
  - Let us define A_i as the event that the i th chosen unit is not defective for i = 1, 2, 3. We are interested in P$(A_1 \cap A_2 \cap A_3)$

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)\,P(A_2|A_1)\,P(A_3|A_2 \cap A_1) \\ &= \frac{95}{100}\frac{94}{99}\frac{93}{98} \\ &\boxed{= 0.8560} \end{aligned}$$

Independence of events

Independence Derivation

- If events A and B satisfy

$$P(A|B) = P(A|B^c) \quad (1)$$

then we say that ***A does not depend on B***

- We can rewrite (1) as

$$\frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B^c)}{P(B^c)}$$

- Using

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

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

Independence Derivation (2)

- We get

$$\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A) - \mathrm{P}(A \cap B)}{1 - \mathrm{P}(B)}$$

$$\mathrm{P}(A \cap B)\big(1 - \mathrm{P}(B)\big) = \mathrm{P}(B)(\mathrm{P}(A) - \mathrm{P}(A \cap B))$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

Independence Derivation (3)

- The above sequence of calculation is reversible

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \Rightarrow \mathrm{P}(A|B) \ = \ \mathrm{P}(A|B^c)$$

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

$$\mathrm{P}(A|B^c) = \frac{\mathrm{P}(A \cap B^c)}{\mathrm{P}(B^c)} = \frac{\mathrm{P}(A) - \mathrm{P}(A \cap B)}{1 - \mathrm{P}(B)} = \frac{\mathrm{P}(A) - \mathrm{P}(A)\mathrm{P}(B)}{1 - \mathrm{P}(B)} = \mathrm{P}(A)$$

$$\Rightarrow \mathrm{P}(A|B) \ = \ \mathrm{P}(A|B^c)$$

Independence Derivation (4)

- Interchanging the role of A and B in the above equations, we get that if B does not depend on A then
$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$
- Since sequence of calculation is reversible and $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$ is symmetric in A and B it follows that
- A does not depend on B iff B does not depend on A

Definition of Independence

- Events A and B are said to be statistically independent or simply independent if their joint probability equals the product of their probabilities

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

- Equivalently:
  - $P(A) = 0$, or $P(A) \neq 0$ & $P(B|A) = P(B)$
  - $P(B) = 0$, or $P(B) \neq 0$ & $P(A|B) = P(A)$

- Interpretation: A and B independent means that the occurrence of one tells nothing about the occurrence of the other.

Question

- Show that events A and B are independent if $\mathrm{P}(B) = 0$

- Solution: Suppose that A and B are any two events. If P(B) = 0, then we claim that A and B are independent. We must show that $P(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) = 0$.
- To show that the left-hand side is zero, observe that since probabilities are nonnegative, and since $A \cap B \subset B$,

$$0 \leq \mathrm{P}(A \cap B) \leq \mathrm{P}(B) = 0$$

Independence

- If the events A and B are independent, are
  - A and B^c
  - A^c and B
  - A^c and B^c

  independent?

Proof

We prove A and B^c are independent.

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

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

$$\Longrightarrow P(A \cap B^c) = P(A) - P(A)\,P(B) \quad (A, B \text{ independent})$$

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

Similarly, one can easily prove A^c and B are independent and A^c and B^c are independent.

Example

- toss 3 coins

$$\Omega = \{HHH, HHT, HTH, \ldots, TTT\},$$
$$\mathcal{A} = 2^{\Omega}, P(A) = |A|/8$$

  - A = even numbers of H's and B = first toss of H
  - Are A and B independent?

$$A = \{HHT, HTH, THH, TTT\}, \quad B = \{HHH, HHT, HTH, HTT\}$$

$$\mathrm{P}(A) = \frac{4}{8}, \mathrm{P}(B) = \frac{4}{8}$$

$$A \cap B = \{HHT, HTH\}, \quad \Rightarrow \mathrm{P}(A \cap B) = \frac{2}{8}$$

$$so\ \mathrm{P}(A)\,\mathrm{P}(B) = \mathrm{P}(A \cap B)$$

Mutual Independence Definition

- Events $A_1, A_2, \ldots, A_n$ are mutual independent if the following relation holds

$$P\left(A_{k_1} \cap A_{k_2} \ldots \cap A_{k_l}\right) = \prod_{i=1}^{l} P\left(A_{k_i}\right)$$

for any subset $\{k_1, k_2, \ldots, k_l\}$ of {1, 2, . . . , n}.

  – The above definition implies that any set of two events, three events, and so on, must satisfies this equation.

Important Remarks

- Pairwise independence does not imply mutual independence

- Example: Let $\Omega = \{1,2,3,4,5,6,7\}, \mathcal{A} = 2^{\Omega}$

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{8}, P(\{7\}) = \frac{1}{4}$$

Consider

$$A = \{1,2,7\}, B = \{3,4,7\}, C = \{5,6,7\}$$

Important Remarks

- Note that $A, B, C \in \mathcal{A}$ and
$$P(A) = P(B) = P(C) = 1/2$$
- Note that
$$P(A \cap B) = P(\{7\}) = \frac{1}{4}$$
$$P(A \cap C) = P(\{7\}) = \frac{1}{4}$$
$$P(B \cap C) = P(\{7\}) = \frac{1}{4}$$

Important Remarks

- Then

$$\frac{1}{4} = P(A \cap B) = P(A)P(B) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$
$$\frac{1}{4} = P(A \cap C) = P(A)P(C) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$
$$\frac{1}{4} = P(B \cap C) = P(B)P(C) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

So, A,B,C are pairwise independent.

- But $P(A \cap B \cap C) = \mathrm{P}(\{7\}) = \frac{1}{4}, P(A)P(B)P(C) = \frac{1}{8}$

- So $P(A \cap B \cap C) \neq P(A)P(B)P(C)$, A, B, C are not mutually independent

Important Remarks(2)

- Three events A, B, C such that $P(A \cap B \cap C) = P(A)P(B)P(C)$ does not imply A, B, C are pairwise independent.

- Example: let $\Omega = \{1,2,3,4\}, \mathcal{A} = 2^{\Omega}$ and

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = p = \frac{3-\sqrt{3}}{4}$$

$$P(\{4\}) = q = \frac{-5+3\sqrt{3}}{4}$$

Let $A = \{1,4\}, B = \{2,4\}, C = \{3,4\}$

Important Remarks(2)

- *Note that as required*

$$3p + q = \frac{9 - 3\sqrt{3} - 5 + 3\sqrt{3}}{4} = 1$$

- *further*

$$(p+q)^3 = q, (p+q)^2 \neq q$$

- *Then*

$$A \cap B \cap C = \{4\}, \quad A \cap B = B \cap C = A \cap C = \{4\}$$

Important Remarks(2)

$$P(A \cap B \cap C) = P(\{4\}) = q$$

$$P(A) = P(B) = P(C) = p + q$$

$$P(A)\,P(B)\,P(C) = (p+q)^3 = q$$

$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = q$$

$$P(A)\,P(B) = P(B)\,P(C) = P(A)\,P(C) = (p+q)^2 \neq q$$

So A, B, C are not pairwise independent.

Important Remarks: Disjoint and Independent

- Difference between disjoint events and independent events
- Recall that events A and B are disjoint if $A \cap B = \emptyset$
  - This concept does not involve the probability measure P in any way
  - To determine if A and B are disjoint only knowledge of A and B is required
  - The notation of independence of events A and B does involve the probability measure P
    - Note that $P(A \cap B) = P(A)P(B)$ does depend on P and not just on A and B

Thank You!