MATH221
Mathematics for Computer Science

Unit 11
Probability

Learning Objectives

- Fundamentals knowledge of probability
- Conditional Probability
- Bayes' Theorem
- Independent Events

1: Probability Axioms

Probability Axioms

Probability Axioms

Recall: a sample space is a set of all outcomes of a random process or experiment and that an event is a subset of a sample space.

Probability Axioms

Let S be a sample space. A probability function P from the set of all events in S to the set of real numbers satisfies the following axioms: For all events A and B in S,

1. $0 \leq P(A) \leq 1$
2. $P(\emptyset) = 0$ and $P(S) = 1$
3. If A and B are disjoint ($A \cap B = \emptyset$), then $P(A \cup B) = P(A) + P(B)$

Probability Axioms

Probability of the Complement of an Event

Suppose that A is an event in a sample space S. Deduce that $P(A^c) = 1 - P(A)$.

By Complement Laws:

(a) $A \cup A^c = U$ and (b) $A \cap A^c = \varnothing$

With S playing the role of the universal set U,

$A \cup A^c = S$ and $A \cap A^c = \varnothing$

Thus S is the disjoint union of A and A^c, and so

$P(A \cup A^c) = P(A) + P(A^c) = P(S) = 1$

Subtracting $P(A)$ from both sides:

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

Probability Axioms

Probability of the Complement of an Event

Probability of the Complement of an Event

If A is any event in a sample space S, then

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

Probability Axioms

Probability of a General Union of Two Events

Probability of a General Union of Two Events

If A and B are any events in a sample space S, then

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Note: This unit shall focus on the use of all these basic properties. However, their proofs can be fund in supplementary notes.

2: Conditional Probability, Bayes' Theorem, and Independent Events

Conditional Probability

Conditional Probability

Imagine a couple with two children, each of whom is equally likely to be a boy or a girl. Now suppose you are given the information that one is a boy. What is the probability that the other child is a boy?

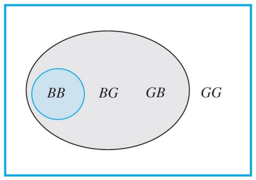
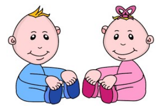


Figure 11.11.1



New sample space = gray region.

Conditional Probability

Within the new sample space, there is one combination where the other child is a boy (blue-gray region).

Hence, the likelihood that the other child is a boy given that at least one is a boy = 1/3.

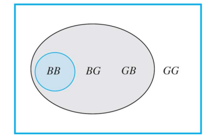


Figure 11.8.1

Note also

$$\frac{P(\text{at least one child is a boy and the other child is also a boy})}{P(\text{at least one child is a boy})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Conditional Probability

A generalization of this observation forms the basis for the following definition.

Definition: Conditional Probability

Let A and B be events in a sample space S. If $P(A) \neq 0$, then the **conditional probability of B given A**, denoted $P(B|A)$, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad 11.8.1$$

Multiplying both sides of formula 11.8.1 by $P(A)$, we

$$P(A \cap B) = P(B|A) \cdot P(A) \quad 11.8.2$$

Dividing both sides of formula 11.8.1 by $P(B|A)$, we get

$$P(A) = \frac{P(A \cap B)}{P(B|A)} \quad 11.8.3$$

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

An urn contains 5 blue and 7 gray balls. Let us say that 2 are chosen at random, one after the other, without replacement

a. Find the following probabilities and illustrate them with a tree diagram: the probability that both balls are blue, the probability that the first ball is blue and the second is not blue, the probability that the first ball is not blue and the second ball is blue, and the probability that neither ball is blue.

b. What is the probability that the second ball is blue?

c. What is the probability that at least one of the balls is blue?

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

An urn contains 5 blue and 7 gray balls. Let us say that 2 are chosen at random, one after the other, without replacement

Let

- S denote the sample space of all possible choices of two balls from the urn,
- B_1 be the event that the first ball is blue (then $B_1{}^c$ is the event that the first ball is not blue),
- B_2 be the event that the second ball is blue (then $B_2{}^c$ is the event that the second ball is not blue).

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

An urn contains 5 blue and 7 gray balls. Let us say that 2 are chosen at random, one after the other, without replacement.

a. Find the following probabilities and illustrate them with a tree diagram: the probability that both balls are blue, the probability that the first ball is blue and the second is not blue, the probability that the first ball is not blue and the second ball is blue, and the probability that neither ball is blue.

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

5 blue, 7 gray. a. Find $P(B_1 \cap B_2)$, $P(B_1 \cap B_2{}^c)$, $P(B_1{}^c \cap B_2)$, $P(B_1{}^c \cap B_2{}^c)$.

$P(B_1) = \frac{5}{12}$ $\quad P(B_1^c) = \frac{7}{12}$ $\quad P(B_2|B_1) = \frac{4}{11}$ $\quad P(B_2^c|B_1) = \frac{7}{11}$

$P(B_2|B_1^c) = \frac{5}{11}$ $\quad P(B_2^c|B_1^c) = \frac{6}{11}$

By formula 10.8.2

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

$$P(B_1 \cap B_2) = P(B_2|B_1) \cdot P(B_1) = \frac{4}{11} \cdot \frac{5}{12} = \frac{20}{132}$$

$$P(B_1 \cap B_2^c) = P(B_2^c|B_1) \cdot P(B_1) = \frac{7}{11} \cdot \frac{5}{12} = \frac{35}{132}$$

$$P(B_1^c \cap B_2) = P(B_2|B_1^c) \cdot P(B_1^c) = \frac{5}{11} \cdot \frac{7}{12} = \frac{35}{132}$$

$$P(B_1^c \cap B_2^c) = P(B_2^c|B_1^c) \cdot P(B_1^c) = \frac{6}{11} \cdot \frac{7}{12} = \frac{42}{132}$$

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

5 blue, 7 gray. a. Find $P(B_1 \cap B_2)$, $P(B_1 \cap B_2{}^c)$, $P(B_1{}^c \cap B_2)$, $P(B_1{}^c \cap B_2{}^c)$.

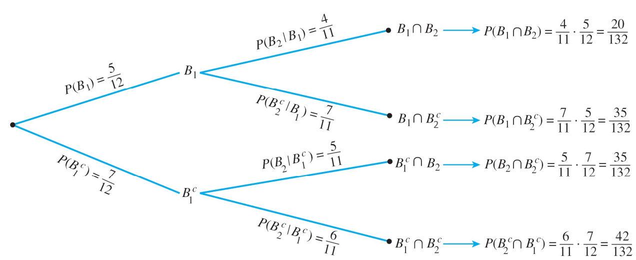


Figure 11.8.2

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

5 blue, 7 gray. b. What is the probability that the 2nd ball is blue?

First ball is blue and second is also blue, OR
first ball is gray and second is blue.
They are mutually exclusive.

$$P(B_2) = P((B_2 \cap B_1) \cup (B_2 \cap B_1^c))$$

$$= P(B_2 \cap B_1) + P(B_2 \cap B_1^c) \quad \text{by probability axiom 3}$$

$$= \frac{20}{132} + \frac{35}{132} = \frac{55}{132} = \mathbf{\frac{5}{12}}$$

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

5 blue, 7 gray. c. What is the probability that at least one ball is blue?

Probability of a General Union of Two Events

If A and B are any events in a sample space S, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$$

$$= \frac{5}{12} + \frac{5}{12} - \frac{20}{132} = \frac{90}{132} = \mathbf{\frac{15}{22}}$$

Conditional Probability

Example 1 – Representing Conditional Probabilities in a Tree Diagram

5 blue, 7 gray. d. What are the probabilities of no blue ball, one blue ball and two blue balls?

$P(\text{no blue balls}) = 1 - P(\text{at least one blue ball}) = 1 - \frac{15}{22} = \frac{7}{22}$

The event that one ball is blue → first ball is blue but second ball is not; or second ball is blue but first ball is not.

From part (a), $P(B_1 \cap B_2^c) = \frac{35}{132}$ and $P(B_1^c \cap B_2) = \frac{35}{132}$

Hence, $P(\text{1 blue ball}) = \frac{35}{132} + \frac{35}{132} = \frac{70}{132}$

From part (a), $P(\text{2 blue balls}) = \frac{20}{132}$

Bayes' Theorem

Bayes' Theorem

Suppose that one urn contains 3 blue and 4 gray balls and a second urn contains 5 blue and 3 gray balls. A ball is selected by choosing one of the urns at random and then picking a ball at random from that urn. If the chosen ball is blue, what is the **probability that it came from the first urn**?

This problem can be solved by carefully interpreting all the information that is known and putting it together in just the right way.

Let

- A be the event that the chosen ball is blue,
- B_1 the event that the ball came from the first urn, and
- B_2 the event that the ball came from the second urn.

Bayes' Theorem

Bayes' Theorem

3 of the 7 balls in the first urn are blue, 5 of the 8 balls in the second urn are blue: $P(A|B_1) = \frac{3}{7}$ and $P(A|B_2) = \frac{5}{8}$

The urns are equally like to be chosen: $P(B_1) = P(B_2) = \frac{1}{2}$

By formula 10.8.2 $P(A \cap B) = P(B|A) \cdot P(A)$

$$P(A \cap B_1) = P(A|B_1) \cdot P(B_1) = \frac{3}{7} \cdot \frac{1}{2} = \frac{3}{14}$$

$$P(A \cap B_2) = P(A|B_2) \cdot P(B_2) = \frac{5}{8} \cdot \frac{1}{2} = \frac{5}{16}$$

A is a disjoint union of $(A \cap B_1)$ and $(A \cap B_2)$, so by probability axiom 3,

$$P(A) = P\big((A \cap B_1) \cup (A \cap B_2)\big)$$
$$= P(A \cap B_1) + P(A \cap B_2) = \frac{3}{14} + \frac{5}{16} = \frac{59}{112}$$

Bayes' Theorem

Bayes' Theorem

3 of the 7 balls in the first urn are blue, 5 of the 8 balls in the second urn are blue.

From previous slide, $P(A \cap B_1) = \frac{3}{14}$ and $P(A) = \frac{59}{112}$

By definition of conditional probability,

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{\frac{3}{14}}{\frac{59}{112}} = \frac{336}{826} \cong 40.7\%$$

Thus, if the chosen ball is blue, the probability that it came from the first urn is approximately 40.7%.

Bayes' Theorem

Theorem 10.8.1 Bayes' Theorem

Suppose that a sample space S is a union of mutually disjoint events $B_1, B_2, B_3, \ldots, B_n$.

Suppose A is an event in S, and suppose A and all the B_i have non-zero probabilities.

If k is an integer with $1 \le k \le n$, then

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \cdots + P(A|B_n) \cdot P(B_n)}$$

Bayes' Theorem

Example 2 – Applying Bayes' Theorem

Most medical tests occasionally produce incorrect results, called false positives and false negatives.

When a test is designed to determine whether a patient has a certain disease:

A **true positive** result indicates that a patient has the disease when the patient has it.

A **false positive** result indicates that a patient has the disease when the patient does not have it.

A **true negative** result indicates that a patient does not have the disease when the patient does not have it.

A **false negative** result indicates that a patient does not have the disease when the patient does have it.

Bayes' Theorem

Example 2 – Applying Bayes' Theorem

Consider a medical test that screens for a disease found in 5 people in 1,000. Suppose that the false positive rate is 3% and the false negative rate is 1%.

That is, 3% of the time a person who does not has the condition tests positive for it, and 1% of the time a person who has the condition tests negative for it.

a. What is the probability that a randomly chosen person who tests positive for the disease actually has the disease?

b. What is the probability that a randomly chosen person who tests negative for the disease does not indeed have the disease?

Bayes' Theorem

Example 2 – Applying Bayes' Theorem

Consider a person chosen at random from among those screened. Let

- A be the event that the person tests positive for the disease,
- B_1 the event that the person actually has the disease, and
- B_2 the event that the person does not have the disease.

Then, $P(A|B_1) = 1 - P(A^c|B_1)$ and $P(A^c|B_2) = 1 - P(A|B_2)$

Hence, $P(A^c|B_1) = 0.01$ $\quad P(A|B_2) = 0.03$

$P(A|B_1) = 0.99$ $\quad P(A^c|B_2) = 0.97$

Also, because 5 people in 1000 have the disease,

$$P(B_1) = 0.005 \qquad P(B_2) = 0.995$$

Bayes' Theorem

Example 2 – Applying Bayes' Theorem

a. What is the probability that a randomly chosen person who tests positive for the disease actually has the disease?

By Bayes' Theorem,

$$P(B_1|A) = \frac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)}$$

$$= \frac{(0.99) \cdot (0.005)}{(0.99) \cdot (0.005) + (0.03) \cdot (0.995)}$$

$$\cong 0.1422 \cong \mathbf{14.2\%}$$

Thus the probability that a person with a positive test result actually has the disease is approximately 14.2%.

Bayes' Theorem

Example 2 – Applying Bayes' Theorem

b. What is the probability that a randomly chosen person who tests negative for the disease does not indeed have the disease?

By Bayes' Theorem,

$$P(B_2|A^c) = \frac{P(A^c|B_2) \cdot P(B_2)}{P(A^c|B_1) \cdot P(B_1) + P(A^c|B_2) \cdot P(B_2)}$$

$$= \frac{(0.97) \cdot (0.995)}{(0.01) \cdot (0.005) + (0.97) \cdot (0.995)}$$

$$\cong 0.999948 \cong \mathbf{99.995\%}$$

Thus the probability that a person with a negative test result does not have the disease is approximately 99.995%.

Independent Events

For convenience and to eliminate the requirement that the probabilities be nonzero, we use the following product formula to define independent events.

Definition: Independent Events

If A and B are events in a sample space S, then A and B are **independent**, if and only if,

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

Independent Events

Example 3 – Computing Probabilities of Intersections of Two Independent Events

A coin is loaded so that the probability of heads is 0.6. Suppose the coin is tossed twice. Although the probability of heads is greater than the probability of tails, there is no reason to believe that whether the coin lands heads or tails on one toss will affect whether it lands heads or tails on the other toss. Thus it is reasonable to assume that the results of the tosses are independent.

a. What is the probability of obtaining two heads?

b. What is the probability of obtaining one head?

c. What is the probability of obtaining no heads?

d. What is the probability of obtaining at least one head?

Independent Events

Example 3 – Computing Probabilities of Intersections of Two Independent Events

Sample space S consists of the 4 outcomes {HH, HT, TH, TT} which are not equally likely.

Let

- E be the event that a head is obtained on the first toss
- F be the event that a head is obtained on the second toss

$P(E) = P(F) = 0.6$.

Independent Events

Example 3 – Computing Probabilities of Intersections of Two Independent Events

a. What is the probability of obtaining two heads?

$P(\text{two heads}) = P(E \cap F) = P(E) \cdot P(F) = (0.6)(0.6) = 0.36 =$ **36%**

b. What is the probability of obtaining one head?

$P(\text{one head}) = P((E \cap F^c) \cup (E^c \cap F)) = P(E) \cdot P(F^c) + P(E^c) \cdot P(F)$

$= (0.6)(0.4) + (0.4)(0.6) = 0.48 =$ **48%**

c. What is the probability of obtaining no heads?

$P(\text{no heads}) = P(E^c \cap F^c) = P(E^c) \cdot P(F^c) = (0.4)(0.4) = 0.16 =$ **16%**

Independent Events

Example 3 – Computing Probabilities of Intersections of Two Independent Events

d. What is the probability of obtaining at least one head?

Method 1:

P(at least one head) = P(one head) + P(two heads) = 0.48 + 0.36
= 0.84 = **84%**

Method 2:

P(at least one head) = 1 – P(no heads) = 1 – 0.16 = 0.84 = **84%**

Pairwise Independent/Mutually Independent

Pairwise Independent/Mutually Independent

We say three events A, B, and C are *pairwise independent* if, and only if,

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

$P(B \cap C) = P(B) \cdot P(C)$

Events can be pairwise independent without satisfying the condition
$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$.

Conversely, they can satisfy the condition
$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ without being pairwise independent.

Pairwise Independent/Mutually Independent

Four conditions must be included in the definition of independence for three events

Definition: Pairwise Independent and Mutually Independent

Let A, B and C be events in a sample space S. A , B and C are **pairwise independent**, if and only if, they satisfy conditions 1 – 3 below. They are **mutually independent** if, and only if, they satisfy all four conditions below.

1. $P(A \cap B) = P(A) \cdot P(B)$
2. $P(A \cap C) = P(A) \cdot P(C)$
3. $P(B \cap C) = P(B) \cdot P(C)$
4. $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

Pairwise Independent/Mutually Independent

The definition of mutual independence for any collection of n events with $n \geq 2$ generalizes the two definitions given previously.

Definition: Mutually Independent

Events $A_1, A_2, \ldots, A_n$ in a sample space S are **mutually independent** if, and only if, the probability of the intersection of any subset of the events is the product of the probabilities of the events in the subset.

Pairwise Independent/Mutually Independent

Example 4 – Tossing a Loaded Coin Ten Times

A coin is loaded so that the probability of heads is 0.6 (and thus the probability of tails is 0.4). Suppose the coin is tossed ten times. As in Example 3, it is reasonable to assume that the results of the tosses are mutually independent.

a. What is the probability of obtaining eight heads?

b. What is the probability of obtaining at least eight head?

For each i = 1, 2, . . . , 10, let H_i be the event that a head is obtained on the i-th toss, and let T_i be the event that a tail is obtained on the i-th toss.

Pairwise Independent/Mutually Independent

Example 4 – Tossing a Loaded Coin Ten Times

a. What is the probability of obtaining eight heads?

Suppose that the eight heads occur on the first eight tosses and that the remaining two tosses are tails. This is the event

$$H_1 \cap H_2 \cap H_3 \cap H_4 \cap H_5 \cap H_6 \cap H_7 \cap H_8 \cap T_9 \cap T_{10}.$$

For simplicity, we denote it as *HHHHHHHHTT*.

By definition of mutually independent events,

$$P(HHHHHHHHTT) = (0.6)^8(0.4)^2$$

By commutative law for multiplication, if the eight heads occur on any other of the ten tosses, the same number is obtained. Eg:

$$P(HHTHHHHHTH) = (0.6)^2(0.4)(0.6)^5(0.4)(0.6) = (0.6)^8(0.4)^2$$

Pairwise Independent/Mutually Independent

Example 4 – Tossing a Loaded Coin Ten Times

a. What is the probability of obtaining eight heads?

Now there are as many different ways to obtain eight heads in ten tosses as there are subsets of eight elements (the toss numbers on which heads are obtained) that can be chosen from a set of ten elements. This number is $\binom{10}{8}$.

Hence

$$P(\text{eight heads}) = \binom{10}{8}(0.6)^8(0.4)^2$$

Pairwise Independent/Mutually Independent

Example 4 – Tossing a Loaded Coin Ten Times

b. What is the probability of obtaining at least eight heads?

By similar reasoning,

$$P(\text{nine heads}) = \binom{10}{9}(0.6)^9(0.4)$$

and

$$P(\text{ten heads}) = \binom{10}{10}(0.6)^{10}$$

Therefore,

$$P(\text{at least 8 heads}) = P(\text{8 heads}) + P(\text{9 heads}) + P(\text{10 heads})$$

$$= \binom{10}{8}(0.6)^8(0.4)^2 + \binom{10}{9}(0.6)^9(0.4) + \binom{10}{10}(0.6)^{10}$$

$\cong$ 0.167 = **16.7%**

End of Unit 11