Lecture 9

188 200
Discrete Mathematics and Linear Algebra

Pattarawit Polpinit

Department of Computer Engineering
Khon Kaen Uiversity

July 13, 2009

Overview

Topic for today.

- Binomial theorem
- Probability axioms
- Expected value
- Conditional probability
- Bayes' theorem and its applications

Reference : Section 6.7-6.9

The Binomial Theorem

In algebra, a sum of two term, such as $a + b$, is call a **binomial**.

The **binomial theorem** gives an expression for the powers of a binomial $(a + b)^n$, $n \geq 0$ and $a, b \in \mathbb{R}$.

Binomial Theorem: Given $a, b \in \mathbb{R}$ and integer $n \geq 0$.

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

The Binomial Theorem: Example

Example: Expand $(a+b)^5$.

Solution: Using the binomial theorem, we compute

$$\begin{aligned}(a+b)^5 &= \sum_{k=0}^{5} \binom{5}{k} a^{5-k} b^k \\ &= \binom{5}{0} a^5 b^0 + \binom{5}{1} a^4 b^1 + \binom{5}{2} a^3 b^2 + \binom{5}{3} a^2 b^3 \\ &\quad + \binom{5}{2} a^1 b^4 + \binom{5}{1} a^0 b^5 \\ &= a^5 + 5a^4 b + 10a^3 b^2 + 10a^2 b^3 + 5ab^4 + b^5\end{aligned}$$

Theory of Probability

So far, we have learned "basic of counting", "permutation", "combination", and a little bit of theory of probability.

Theory of probability was originated from studying gambling, e.g. probability of a die comes up odd number.

There are many applications for probability theory:

- risk assessment
- genetics
- simulation
- algorithm design
- gambling
- reliability

Probability Axioms

Probability Axioms:

Experiment is a procedure that produces a possible outcome.

Let S be a sample space, A and B be any events in S then.

- Recall that probability of an event A is $P(A) = \frac{|A|}{|S|}$.
- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$ and $P(S) = 1$.
- Probability of the complement of an event:

$$P(\overline{A}) = 1 - P(A)$$

- Probability of a union of two events:

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

Probability Axioms: Example

Example: Suppose a card is chosen at random from a deck of card. What is the probability that the card is a face card or is from one of the red suit?

Let's review a deck of card:

- There are 52 cards of which are dived into 4 suits.
- Red suit: diamonds (♢) and hearts (♡), **black suit:** clubs (♣) and spades (♠).
- Each suite contains 13 cards: 2, 3,. . ., 10, J, Q, K, A.
- The face cards are J, Q and K.

Solution:

Expected Value

Many questions can be formulated in terms of the value we expect a random variable to take.

- or more precisely, the average value of a random variable of a large number of experiments.
- E.g. how many heads are expected after a coin is flipped 100 times?

Suppose the possible outcomes T of an experiment are a_1, a_2, ..., $a_n \in \mathbb{R}$ which occur with probabilities p_1, p_2, ..., p_n. **The expected value ($E(T)$)** is

$$\sum_{k=1}^{n} a_k p_k = a_1 p_1 + a_2 p_2 + \ldots + a_n p_n$$

Expected Value: Example

Example: Let T be the number that comes up when a dice is rolled. What is the expected value of T?

Solution:

- There are 6 possible outcomes: $S = \{1, 2, 3, 4, 5, 6\}$.
- Each out come occurs with the same probability $1/6$.
- Hence, the expected value of T is

$$\begin{aligned} E(T) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{21}{6} \\ &= \frac{7}{2} \end{aligned}$$

Expected Value: Example 2

Example: We flip three coins, one at a time. Let S be the sample space. Let T be the possible outcome that coins turn head. What is the expected value of T?

Solution:

- There are 8 possible outcomes: $\{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$
- Since the coin are fair, the probability of each outcome is $1/8$.
- Let $a_1, a_2, \ldots, a_8$ be the number of heads that appear when HHH, HHT, ..., TTT are the outcomes.
  - Hence a_1, the number of head that appear when the outcome is HHH, is 3.
  - $a_2 = a_3 = a_5 = 2$.
  - $a_4 = a_6 = a_7 = 1$.
  - $a_8 = 0$.

Expected Value: Example 2 cont.

Therefore,

$$\begin{aligned} E(T) &= \frac{1}{8}a_1 + \frac{1}{8}a_2 + \ldots + \frac{1}{8}a_8 \\ &= \frac{1}{8}(a_1 + a_2 + \ldots + a_8) \\ &= \frac{1}{8}(3 + 2 + 2 + 1 + 2 + 1 + 1 + 1) \\ &= \frac{12}{8} \\ &= \frac{3}{2} \end{aligned}$$

In other words, if we keep flipping the coins many times, the average that coins will turn head will be 1.5.

Expected Value: Alternating Version

If the experiment has relatively few outcomes, we can compute the expected value directly from its definition.

However, when an experiment has a large number of outcome, it may be inconvenient to use its original definition. We can group together all outcomes assigned the same random value.

Suppose T is a random variable with range $\{r_1, r_2, \ldots, r_m\}$ and let $p(T = r_k)$ be the probability that the random variable T takes r_k. Then the expected value of T is

$$E(T) = \sum_{k=1}^{m} p(T = r_k) r_k$$

Let's see an example of this.

Expected Value: Example 3

Example: What is the expected value of the sum of the number that appear when a pair of dice is rolled?

Solution:

- Let (x, y) be an outcome when a pair of dice is rolled if x appears on the first die and y is what appears on the second die. There are 36 possible outcomes and their sum are.
  - $(1,1) = 2$
  - $(1,2) = (2,1) = 3$
  - $(1,3) = (2,2) = (3,1) = 4$
  - $(1,4) = (2,3) = (3,2) = (4,1) = 5$
  - $(1,5) = (2,4) = (3,3) = (4,2) = (5,1) = 6$
  - $(1,6) = (2,5) = (3,4) = (4,3) = (5,2) = (1,6) = 7$
  - $(2,6) = (3,5) = (4,4) = (5,3) = (6,2) = 8$
  - $(3,6) = (4,5) = (5,4) = (6,3) = 9$
  - $(4,6) = (5,5) = (6,4) = 10$
  - $(5,6) = (6,5) = 11$
  - $(6,6) = 12$

Expected Value: Example 3 cont.

- Hence the possible outcomes of the sum (the range) are $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.
- Let T be a possible value of the sum of two dice and $p(T = r)$ be the probability when the sum is r. The probabilities of each outcomes of the sum are
  - $p(T = 2) = p(T = 12) = 1/36$
  - $p(T = 3) = p(T = 11) = 2/36$
  - $p(T = 4) = p(T = 10) = 3/36$
  - $p(T = 5) = p(T = 9) = 4/36$
  - $p(T = 6) = p(T = 8) = 5/36$
  - $p(T = 7) = 6/36$
- Therefore, the expected value of the sum of two dice is

$$\begin{aligned} E(T) &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + 5 \cdot \frac{1}{9} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{1}{6} \\ &\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{1}{9} + 10 \cdot \frac{1}{12} + 11 \cdot \frac{1}{18} + 12 \cdot \frac{1}{36} \\ &= 7 \end{aligned}$$

Expected Value: Example 2 cont.

What do you do if you want to use its original expected value definition?

- There are 36 possible outcomes.
- For each outcome, we compute the sum of two dice.
- The expected value would be

$$\begin{aligned} E(T) &= (1,1) \cdot \frac{1}{36} + (1,2) \cdot \frac{1}{36} + (1,3) \cdot \frac{1}{36} + \ldots + (6,6) \cdot \frac{1}{36} \\ &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{36} + 4 \cdot \frac{1}{36} + \ldots + 12 \cdot \frac{1}{36} \end{aligned}$$

Expected Value of a Lottery

Example: Suppose that the Thai government lottery office sells 1,000,000 lotteries for 40 baht each. Assume that half the lotteries are sold. The prizes are as follows

- The first prize is 2,000,000 baht
- 5 second prizes for 100,000 baht each
- 10 third prizes for 40,000 baht each
- 50 fourth prizes for 20,000 each
- 100 fifth prizes for 10,000 each
- 2 next-to-first-prize prizes for 50,000 each
  - If the first prize is 123456 the next-to-first-prize prizes are 123455 and 123457.

What is the expected value when you buy a lottery?

Expected Value of a Lottery cont.

Solution:

- Let p_1 be the probability of a lottery being the first prize, p_2, p_3, p_4, p_5, p_6 be the probability of a lotteries being the second prize, and so on ....
  - Each p_k for $1 \leq k \leq 500,000$ has the same probability of 1/500,000.
- Let a_1 be the net profit to buyer who won the first prize, a_2, a_3, a_4, a_5, a_6 be the net profits to buyers who won the second prizes, and so on....
  - E.g., $a_1 = 1,999,960$ (2,000,000 - 40)
- There are 168 winning lotteries in total.
  - which means 499832 lotteries do not win.
  - This implies that $a_{169} = a_{170} = \ldots, = a_{500,000} = -40$.

Expected Value of a Lottery cont. II

Therefore, the expected value of buying a lottery is

$$\begin{aligned}
\sum_{k=1}^{500,000} a_k \cdot p_k &= \sum_{k=1}^{500,000} a_k \cdot \frac{1}{500,000} \\
&= \frac{1}{500,000} \sum_{k=1}^{500,000} a_k \\
&= \frac{1}{500,000} (1,999,960 + 5 \cdot 99960 + 10 \cdot 39960 + \\
&\quad 50 \cdot 19960 + 100 \cdot 9960 + 2 \cdot 49960 + 499832(-40)) \\
&= \frac{1}{500,000}(-10500000) \\
&= -21
\end{aligned}$$

In other words, if you play a lottery for a long time, the average return would be -21 baht.

Conditional Probability

Sometimes, we want to know the probability of some events given that another event has occurred.

Example: What is the probability that a family with two children, they will both be boy?

- Sample space: {BB, BG, GB, GG}.
- The probability is 1/4.

What is the probability if we know one of the children is a boy?

- Now the sample space has become: {BB, BG, GB}.
- The probability is 1/3.

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 is

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

Conditional Probability: Example

Example: A jar contains 5 blues and 7 grey balls. We choose two balls one at a time without putting the ball back into the jar.

a. What is the probability that both balls are blue?

Solution to a.: Let E be the event that the first ball chosen from the jar is blue, and F be the event that the second ball chosen from the jar is blue.

The probability that both balls are blue is $P(E \cap F)$. By the conditional probability, $P(E \cap F) = P(F \mid E) \cdot P(E)$. The probability of the first ball is blue is

$$P(E) = \frac{5}{12}$$

If the first ball is blue, then when the second ball is chosen the jar will contain 4 blue and 7 grey balls. Thus $P(F \mid E) = 4/11$. Hence

$$P(E \cap F) = \frac{4}{11} \cdot \frac{5}{12} = \frac{20}{132}$$

Conditional Probability: Example b.

b. What is the probability that the second ball is blue, but the first ball is grey?

Solution to b.:

Conditional Probability: Example c.

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

Solution to c.:

Conditional Probability: Example d.

d. What is the probability that at least one of the ball is blue?

Solution to d.:

Conditional Probability: Example e.

e. If the experiment of choosing 2 balls from the jar were repeated many times, what would be the expected value of the number of blue balls?

Solution to e.:

Bayes' Theorem: Introduction

Bayes' theorem allows us to relate the conditional and marginal probabilities of two random events.

In other words, **Bayes' theorem** will help us asses the probability that an event occurred given only by partial evidence.

Doesn't the formula for conditional probability do this already?

- Conditional probability: $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$
- Yes, but we can't always use the conditional probability formula directly.

Bayes' Theorem: Motivating Example

Example: Suppose that a certain marihuana test correctly identifies a person who use marihuana as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. If a company suspect that 0.5% of its employee use marihuana, what is the probability that an employee that test positive or this drug is actually use marihuana?

Question: Can we use the conditional probability formula to solve the problem?

Let A be an event where x use marihuana.

Let B be an event where x is tested positive for marihuana.

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

Bayes' Theorem Can Help.

In situations like those on the last slide, Bayes' theorem can be applied!

Essentially, the theorem will allow us to calculate $P(A \mid B)$ (the probability that x is a user given his test is positive) assuming that we can derive:

- $P(A)$: Probability that x is a user
- $P(B \mid A)$: Probability that x will test positively given that he is a user.
  - Test success rate
- $P(B \mid \overline{A})$: Probability that x will test positively give that he is not a user.
  - Test false positive rate

It looks like Bayes' theorem can help.

Bayes' Theorem: Example

Example: Suppose that a jar contains 3 blue and 4 grey balls. The second jar contains 5 blue and 3 grey balls. We choose a ball by first randomly select a jar, then pick up a ball from that jar. If the ball is blue, then what is the probability that it came from the first jar?

Solution:

- Let A be the event that the chosen ball is blue.
- Let B be the event that the ball came from the first jar.
- We want to find probability that the chosen ball came from the first jar given that it is blue : $P(B \mid A)$.

Goal: From the conditional probability $P(B \mid A) = P(B \cap A)/P(A)$, can we use what we know to derive $P(B \cap A)$ and $P(A)$?

What do we know?

- Two jars:
  1. First jar: 3 blue, 4 grey balls
  2. Second jar: 5 blue, 3 grey balls
- A jar is selected at random.
- A chosen ball is blue.
- What is the probability that it is from the first jar?

Statement: a jar is selected at random.

- $P(B) = P(\overline{B}) = 1/2$

Statement: the first contains 3 blue and 4 grey

- $P(A|B) = 3/7$

Statement: the second contains 5 blue and 3 grey

- $P(A|\overline{B}) = 5/8$

Applying what we know to solve the problem.

The end goal: Compute $P(B \mid A) = P(B \cap A)/P(A)$.

Note: $P(A \mid B) = P(A \cap B)/P(B)$.

- So $P(A \cap B) = P(A \mid B) \cdot P(B)$.
- We know that $P(A \mid B) = 3/7$ and $P(B) = 1/2$.
- So $P(A \cap B) = (3/7) \cdot (1/2) = 3/14$.

Recall:
$P(B) = P(\overline{B}) = 1/2$
$P(A|B) = 3/7$
$P(A|\overline{B}) = 5/8$

Similarly: $P(A \cap \overline{B}) = P(A \mid \overline{B}) \cdot P(\overline{B}) = (5/8) \cdot (1/2) = 5/16$.

Observation: $A = (A \cap B) \cup (A \cap \overline{B})$. This implies that

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \overline{B}) \\ &= \frac{3}{14} + \frac{5}{16} \\ &= \frac{59}{112} \end{aligned}$$

Compute the End Goal.

The end goal: Compute $P(B \mid A) = P(B \cap A)/P(A)$.

So $P(B \mid A) = (3/14)/(59/112)$

≈ 0.407

Recall:
$P(B) = P(\overline{B}) = 1/2$
$P(A|B) = 3/7$
$P(A|\overline{B}) = 5/8$
$P(A \cap B) = 3/14$
$P(A) = 59/112$

How did we get here?

1. We check what we have from the problem.
2. Rearrange terms to derive $P(B \cap A)$ and $P(A)$.
3. Use the definition of conditional probability to solve the problem.

So what is Bayes' Theorem?

Bayes' Theorem: Suppose A and B are events from some sample space S such that $P(A) \neq 0$ and $P(B) \neq 0$.

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

As we know that $A = (A \cap B) \cup (A \cap \overline{B})$. So $P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$. So an alternative formula form of Bayes' theorem is

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

More generally, if $S = B_1 \cup B_2 \cup \ldots \cup B_n$ where B_i are disjoint and non-zero.

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Why is Bayes' theorem useful?

In a nutshell, the Bayes' theorem is useful if you want to find $P(A|B)$, but you don't know $P(A \cap B)$ and $P(B)$.

In general, it shows how one conditional probability, $P(A|B)$ depends on its inverse $P(B|A)$.

Bayes' Theorem Example 2

Example: Suppose that students at KKU 60% are boys and 40% are girls. The girl students wear trousers or skirts in equal numbers; the boys all wear pants. An observer sees a (random) student from a distance; all they can see is that this student is wearing pants. What is the probability this student is a girl?

Solution: Let's first set up events:

- Let $A =$ "x is wearing pants." $\longrightarrow$ $\overline{A} =$ "x is wearing a skirt."
- Let $B =$ "x is a girl" $\longrightarrow$ $\overline{B} =$ "x is a boy"
- Find $P(B|A)$!

Let's check what we have:

- $P(B) = 0.4$
- $P(\overline{B}) = 0.6$
- $P(A|B) = P(\overline{A}|B) = 0.5$
- $P(A|\overline{B}) = 1$

Bayes' Theorem Example 2 cont.

Compute: $P(B \mid A)$

$$P(B \mid A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

$$= \frac{0.5(0.4)}{0.5(0.4) + 1(0.6)}$$

$$= 1/4$$

Recall:

$P(B) = 0.4$

$P(\overline{B}) = 0.6$

$P(A|B) = P(\overline{A}|B) = 0.5$

$P(A|\overline{B}) = 1$

Conclusion: There is 25% chance that the person seen was a girl, given that they were wearing pants.

Drug Test Example, Revisited

Example: Suppose that a certain marihuana test correctly identifies a person who use marihuana as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. If a company suspect that 0.5% of its employee use marihuana, what is the probability that an employee that test positive for marihuana is actually marihuana user?

Solution: Let's define events:

- Let $A =$ "x uses marihuana"
- So $\overline{A} =$ "x doesn't use marihuana"
- Let $B =$ "x tests positive for marihuana"
- So $\overline{B} =$ "x t est negative for marihuana"

Find $P(A|B)$**!**

Drug Test Example, cont.

Check: values that we are given.

- $P(A) = 0.005$
- $P(\overline{A}) = 0.995$
- $P(B|A) = 0.99$
- $P(B|\overline{A}) = 0.01$

Compute: $P(A|B)$.

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})} \\ &= \frac{0.99(0.005)}{0.99(0.005) + 0.01(0.995)} \\ &= 0.3322 \end{aligned}$$

Conclusion: If an employee tests positive for marihuana use, there is only a 33% chance that they are actually a marihuana user!

Application: Spam filtering

Definition: Spam is unsolicited bulk email. In other words, it's emails that are sent to lots of people who didn't ask for them.

In recent years, spam has become increasingly problematic.

- According to Symantec, a security company, spam now accounts for **90.4%** of all emails.
- This means that **1 out of every 1.1 e-mails is junk**.

To combat this problem, people have developed spam filters based on Bayes' theorem!

How does a Bayesian spam filter work?

Essentially, these filters determine the probability that a message is spam, given that it contains certain keyword.

- Let $A =$ "email x is spam".
- Let $B =$ "email x contains a questionable keyword".

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

From the above equation:

- $P(B|A) =$ Probability that our keyword occurs in spam emails
- $P(B|\overline{A}) =$ Probability that our keyword occurs in emails that are not spam
- $P(A) =$ Probability that an email is spam.
- $P(\overline{A}) =$ Probability that an email is not spam.

Question: How do we compute these parameters?

We can learn these by examining historical email traces

Imagine that we have a database of email messages

We can ask a few intelligent questions to learn the parameters of our Bayesian filter:

- How many of these messages do we consider spam? : $P(A)$
- In the spam emails, how often does our keyword appear? : $P(B|A)$
- In the good emails, how often does our keyword appear? : $P(B|\overline{A})$

Aside: This is what happens every time you click the mark as spam button in your email client!

Given this information, we can apply Bayes theorem!

Filtering spam using a single keyword

Example: Suppose that the keyword "Rolex" occurs in 250 of 2000 known spam emails, and in 5 of 1000 known good emails. Estimate the probability that an incoming message containing the word "Rolex" is spam, assuming that it is equally likely that an incoming message is spam or not spam. If our threshold for classifying a message as spam is 0.9, will we reject this message?

Define events:

- Let A = message is spam
- Let B = message contains the keyword "Rolex"

Gather probabilities from the problem statement.

- $P(A) = P(\overline{A}) = 0.5$
- $P(B|A) = 250/2000 = 0.125$
- $P(B|\overline{A}) = 5/1000 = 0.005$

Filtering spam using a single keyword, cont.

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

Recall:
$P(A) = P(\overline{A}) = 0.5$
$P(B|A) = 0.125$
$P(B|\overline{A}) = 0.005$

Compute $P(A \mid B)$

$$\begin{aligned} P(A \mid B) &= \frac{0.125(0.5)}{0.125(0.5) + 0.005(0.5)} \\ &= 0.962 \end{aligned}$$

Conclusion: Since the probability that our email is spam given that it contains the string "Rolex" is approximately $0.962 > 0.9$, we will flag this email as spam.

Problems with this simple filter

How would you choose a single keyword to use?

- "All natural"
- "Nigeria"
- "Porn"
- ...

Users get upset if false positives occur, i.e., if good emails are incorrectly classified as spam.

How can we fix this?

- Choose keywords such that $P(\text{spam} \mid \text{keyword})$ is very high or very low.
- Filter based on multiple keywords.
  - That is we want to find $P(A|(B_1 \cap B_2 \cap \ldots))$.
  - We will see how after we have learned Independence Events.

Next time

Next time, we will wrap up the probability theory.

- Independent events
- Independence for three events
- Multiple keywords spam filter based on Bayes' theorem.
- Interesting examples