Probability Theory

Section Summary

- Assigning Probabilities
- Probabilities of Complements and Unions of Events
- Conditional Probability
- Independence
- Random Variables

Assigning Probabilities

- Let S be a sample space of an experiment with a finite number of outcomes. We assign a probability $p(s)$ to each outcome s, so that:
  - *i.* $0 \leq p(s) \leq 1$ for each $s \in S$
  - *ii.* $$\sum_{s \in S} p(s) = 1$$
- The function p from the set of all outcomes of the sample space S is called a *probability distribution*.

Probability of an Event

Definition: The probability of the event *E* is the sum of the probabilities of the outcomes in *E*.

$$p(E) = \sum_{s \in E} p(s)$$

Assigning Probabilities

Example: What probabilities should we assign to the outcomes H (heads) and T (tails) when a fair coin is flipped? What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?

Solution: For a fair coin, we have $p(H) = p(T) = ½$.

For a biased coin, we have $p(H) = 2p(T)$.

Because $p(H) + p(T) = 1$, it follows that

$$2p(T) + p(T) = 3p(T) = 1.$$

Hence, $p(T) = 1/3$ and $p(H) = 2/3$.

Uniform Distribution

Definition: Suppose that S is a set with n elements. The *uniform distribution* assigns the probability $1/n$ to each element of S.

Example: Consider again the coin flipping example, but with a fair coin. Now $p(H) = p(T) = 1/2$.

Probability of a Die

Example: Suppose that a die is biased so that 3 appears twice as often as each other number, but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?

Solution: We want the probability of the event $E = \{1,3,5\}$. We have $p(3) = 2/7$ and

$$p(1) = p(2) = p(4) = p(5) = p(6) = 1/7.$$

Hence, $p(E) = p(1) + p(3) + p(5) =$

$$1/7 + 2/7 + 1/7 = 4/7.$$

Probability of Lottery

Example: There are many lotteries that award prizes to people who correctly choose a set of six numbers out of the first n positive integers, where n is usually between 30 and 60. What is the probability that a person picks the correct six numbers out of 40?

Solution: The number of ways to choose six numbers out of 40 is

$$C(40,6) = 40!/(34!6!) = 3{,}838{,}380.$$

Hence, the probability of picking a winning combination is $1/\ 3{,}838{,}380 \approx 0.00000026$.

Probability of a Lottery

Example: What is the probability that the numbers 11, 4, 17, 39, and 23 are drawn in that order from a bin with 50 balls labeled with the numbers 1,2, …, 50 if

a) The ball selected is not returned to the bin.

b) The ball selected is returned to the bin before the next ball is selected.

Solution: Use the product rule in each case.

a) *Sampling without replacement*: The probability is 1/254,251,200 since there are $50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 = 254{,}251{,}200$ ways to choose the five balls.

b) *Sampling with replacement*: The probability is $1/50^5 = 1/312{,}500{,}000$ since $50^5 = 312{,}500{,}000$.

Probabilities of Complements

$$p(\overline{E}) = 1 - p(E)$$

Proof: Since each outcome is in either E or $\overline{E}$, but not both,

$$\sum_{s \in S} p(s) = 1 = p(E) + p(\overline{E}).$$

Probability of Complements

Example: A sequence of 10 bits is chosen randomly. What is the probability that at least one of these bits is 0?

Solution: Let E be the event that at least one of the 10 bits is 0. Then $\overline{E}$ is the event that all of the bits are 1s. The size of the sample space S is 2^{10}. Hence,

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^{10}} = 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

Probability of Union of Events

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Proof: Each outcome s that is in E_1 is included in $p(E_1)$. Each outcome s that is in E_2 is included in $p(E_2)$. If s that is in both E_1 and E_2 then it is included in both $p(E_1)$and $p(E_2)$ but subtracted out in $p(E_1 \cap E_2)$. Thus the left-hand side of the equation equals the right-hand side.

Probability of Unions

Example: What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible by either 2 or 5?

Solution: Let E_1 be the event that the integer is divisible by 2 and E_2 be the event that it is divisible 5? Then the event that the integer is divisible by 2 or 5 is $E_1 \cup E_2$ and $E_1 \cap E_2$ is the event that it is divisible by 2 and 5.

It follows that:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

$$= 50/100 + 20/100 - 10/100 = 3/5.$$

Combinations of Events

Theorem: If E_1, E_2, … is a sequence of pairwise disjoint events in a sample space S, then

$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i)$$

Conditional Probability

Definition: Let E and F be events with $p(F) > 0$. The conditional probability of E given F, denoted by $P(E|F)$, is defined as:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

Example: A bit string of length four is generated at random so that each of the 16 bit strings of length 4 is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

Solution: Let E be the event that the bit string contains at least two consecutive 0s, and F be the event that the first bit is a 0.

- Since $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, $p(E \cap F) = 5/16$.
- Because 8 bit strings of length 4 start with a 0, $p(F) = 8/16 = ½$.

Hence,

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{5/16}{1/2} = \frac{5}{8}.$$

Conditional Probability

Example: What is the conditional probability that a family with two children has two boys, given that they have at least one boy. Assume that each of the possibilities BB, BG, GB, and GG is equally likely where B represents a boy and G represents a girl.

Solution: Let E be the event that the family has two boys and let F be the event that the family has at least one boy. Then $E = \{BB\}$, $F = \{BB, BG, GB\}$, and $E \cap F = \{BB\}$.

- It follows that p(F) = 3/4 and $p(E \cap F) = 1/4$.

Hence,

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Independence

Definition: The events E and F are *independent* if and only if

$$p(E \cap F) = p(E)p(F).$$

Example: Suppose E is the event that a randomly generated bit string of length four begins with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent if the 16 bit strings of length four are equally likely?

Solution: There are eight bit strings of length four that begin with a 1, and eight bit strings of length four that contain an even number of 1s.

- Since the number of bit strings of length 4 is 16,

$$p(E) = p(F) = 8/16 = ½.$$

- Since $E \cap F = \{1111, 1100, 1010, 1001\}$, $p(E \cap F) = 4/16 = 1/4$.

We conclude that E and F are independent, because

$$p(E \cap F) = 1/4 = (½)\,(½) = p(E)\,p(F)$$

Independence

Example: Assume (as in the previous example) that each of the four ways a family can have two children (*BB*, *GG*, *BG*,*GB*) is equally likely. Are the events E, that a family with two children has two boys, and F, that a family with two children has at least one boy, independent?

Solution: Because $E = \{BB\}$, $p(E) = 1/4$. We saw previously that that $p(F) = 3/4$ and $p(E \cap F) = 1/4$. The events E and F are not independent since

$$p(E)\,\mathrm{p}(F) = 3/16 \neq 1/4 = p(E \cap F) .$$

Pairwise and Mutual Independence

Definition: The events E_1, E_2, …, E_n are *pairwise independent* if and only if $p(E_i \cap E_j) = p(E_i)\, \mathrm{p}(E_j)$ for all pairs i and j with $i \leq j \leq n$.

The events are *mutually independent* if

$$p(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = p(E_{i_1})p(E_{i_2}) \cdots p(E_{i_m})$$

whenever i_j, j = 1,2,…., m, are integers with

$$1 \leq i_1 < i_2 < \cdots < i_m \leq n \quad \text{and } m \geq 2.$$

Random Variables

Definition: A *random variable* is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Random Variables

Definition: The *distribution* of a random variable X on a sample space S is the set of pairs $(r, p(X = r))$ for all $r \in X(S)$, where $p(X = r)$ is the probability that X takes the value r.

Example: Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:

$X(HHH) = 3, X(TTT) = 0,$

$X(HHT) = X(HTH) = X(THH) = 2,$

$X(TTH) = X(THT) = X(HTT) = 1.$

Each of the eight possible outcomes has probability 1/8. So, the distribution of $X(t)$ is $p(X = 3) = 1/8$, $p(X = 2) = 3/8$, $p(X = 1) = 3/8$, and $p(X = 0) = 1/8$.

Bayes' Theorem

Section 7.3

Section Summary

- Bayes' Theorem
- Generalized Bayes' Theorem

Motivation for Bayes’ Theorem

- Bayes’ theorem allows us to use probability to answer questions such as the following:
  - Given that someone tests positive for having a particular disease, what is the probability that they actually do have the disease?
  - Given that someone tests negative for the disease, what is the probability, that in fact they do have the disease?
- Bayes’ theorem has applications to medicine, law, artificial intelligence, engineering, and many diverse other areas.

Bayes' Theorem

Bayes' Theorem: Suppose that E and F are events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

Example: We have two boxes. The first box contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects one of the boxes at random. Then he selects a ball from that box at random. If he has a red ball, what is the probability that he selected a ball from the first box.

- Let E be the event that Bob has chosen a red ball and F be the event that Bob has chosen the first box.
- By Bayes' theorem the probability that Bob has picked the first box is:

$$p(F|E) = \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{7/18}{38/63} = \frac{49}{76} \approx 0.645.$$

Derivation of Bayes' Theorem

- Recall the definition of the conditional probability $p(E|\mathrm{F})$:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

- From this definition, it follows that:

$$p(E|F) = \frac{p(E \cap F)}{p(F)} \quad , \quad p(F|E) = \frac{p(E \cap F)}{p(E)}$$

continued →

Derivation of Bayes' Theorem

On the last slide we showed that

$$p(E|F)p(F) = p(E \cap F), \; p(F|E)p(E) = p(E \cap F)$$

Equating the two formulas
for p(E F) shows that

$$p(E|F)p(F) = p(F|E)p(E)$$

Solving for $p(E|F)$ and for $p(F|E)$ tells us that

$$p(E|F) = \frac{p(F|E)p(E)}{p(F)} \;, \; p(F|E) = \frac{p(E|F)p(F)}{p(E)}$$

continued →

Derivation of Bayes' Theorem

On the last slide we showed that:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E)}$$

Note that $$p(E) = p(E|F)p(F) + p(E|\overline{F})p(\overline{F})$$

since $p(E) = p(E \cap F) + p(E \cap \overline{F})$

because $E = E \cap S = E \cap (F \cup \overline{F}) = (E \cap F) \cup (E \cap \overline{F})$

and $(E \cap F) \cap (E \cap \overline{F}) = \emptyset$

By the definition of conditional probability,

$$p(E) = p(E \cap F) + p(E \cap \overline{F}) = p(E|F)p(F) + p(E|\overline{F})p(\overline{F})$$

Hence,

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

Applying Bayes’ Theorem

Example: Suppose that one person in 100,000 has a particular disease. There is a test for the disease that gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result. Find

a) the probability that a person who test positive has the disease.
b) the probability that a person who test negative does not have the disease.

- Should someone who tests positive be worried?

Applying Bayes' Theorem

Solution: Let D be the event that the person has the disease, and E be the event that this person tests positive. We need to compute $p(D|E)$ from $p(\mathrm{D})$, $p(E|D)$, $p(E \mid \overline{D})$, $p(\overline{D})$.

$p(D) = 1/100,000 = 0.00001$ $\quad p(\overline{D}) = 1 - 0.00001 = 0.99999$

$p(E|D) = .99$ $\quad p(\overline{E}|D) = .01$ $\quad p(E|\overline{D}) = .005$ $\quad p(\overline{E}|\overline{D}) = .995$

$$p(D|E) = \frac{p(E|D)p(D)}{p(E|D)p(D) + p(E|\overline{D})p(\overline{D})}$$

$$= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)}$$

$$\approx 0.002$$

Can you use this formula to explain why the resulting probability is surprisingly small?

So, don't worry too much, if your test for this disease comes back positive.

Applying Bayes' Theorem

- What if the result is negative?

$$p(\overline{D}|\overline{E}) = \frac{p(\overline{E}|\overline{D})p(\overline{D})}{p(\overline{E}|\overline{D})p(\overline{D}) + p(\overline{E}|D)p(D)}$$

$$= \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)}$$

$$\approx 0.9999999$$

So, the probability you have the disease if you test negative is
$$p(D|\overline{E}) \approx 1 - 0.9999999 = 0.0000001.$$

- So, it is extremely unlikely you have the disease if you test negative.

Generalized Bayes' Theorem

Generalized Bayes' Theorem: Suppose that E is an event from a sample space S and that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i}^{n} F_i = S$.

Assume that $p(E) \neq 0$ for $i = 1, 2, \ldots, n$. Then

$$p(F_j|E) = \frac{p(E|F_j)p(F_j)}{\sum_{i=1}^{n} p(E|F_i)p(F_i)}.$$

Monty Hall Puzzle

Example: You are asked to select one of the three doors to open. There is a large prize behind one of the doors and if you select that door, you win the prize. After you select a door, the game show host opens one of the other doors (which he knows is not the winning door). The prize is not behind the door and he gives you the opportunity to switch your selection. Should you switch?

Solution: You should switch. We'll prove this using Bayes' Theorem.

Expected Value and Linearity of Expectation

Section Summary

- Expected Value
- Linearity of Expectation
- Applications

Expected Value

Definition: The *expected value* (or *expectation* or *mean*) of the random variable $X(s)$ on the sample space S is equal to

$$E(X) = \sum_{x \in S} p(s)X(s).$$

Example-Expected Value of a Die: Let X be the number that comes up when a fair die is rolled. What is the expected value of X?

Solution: The random variable X takes the values 1, 2, 3, 4, 5, or 6. Each has probability 1/6. It follows that

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}.$$

Expected Value

Theorem 1: If X is a random variable and $p(X = r)$ is the probability that X = r, so that

$$p(X = r) = \sum_{s \in S, X(s) = r} p(s), \text{ then}$$

$$E(X) = \sum_{r \in X(S)} p(X = r)r.$$

Proof: Suppose that X is a random variable with range $X(S)$ and let $p(X = r)$ be the probability that X takes the value r. Consequently, $p(X = r)$ is the sum of the probabilities of the outcomes s such that $X(s) = r$. Hence,

$$E(X) = \sum_{r \in X(S)} p(X = r)r.$$

◀

Linearity of Expectation

The following theorem tells us that expected values are linear. For example, the expected value of the sum of random variables is the sum of their expected values.

Theorem 3: If X_i, $i = 1, 2, \ldots, n$ with n a positive integer, are random variables on S, and if a and b are real numbers, then

$$(i)\ E(X_1 + X_2 + \ldots. + X_n) = E(X_1) + E(X_2) + \ldots. + E(X_n)$$

$$(ii)\ E(aX + b) = aE(X) + b.$$

Proof in Class

Linearity of Expectation

Expected Value in the Hatcheck Problem: A new employee started a job checking hats, but forgot to put the claim check numbers on the hats. So, the n customers just receive a random hat from those remaining. What is the expected number of hat returned correctly?

Solution: Let X be the random variable that equals the number of people who receive the correct hat. Note that $X = X_1 + X_2 + \cdots + X_n$,

where $X_i = 1$ if the ith person receives the hat and $X_i = 0$ otherwise.

- Because it is equally likely that the checker returns any of the hats to the ith person, it follows that the probability that the ith person receives the correct hat is $1/n$. Consequently (by Theorem 1), for all I

$$E(X_i) = 1 \cdot p(X_i = 1) + 0 \cdot p(X_i = 0) = 1 \cdot 1/n + 0 = 1/n .$$

- By the Linearity of Expectation, it follows that:

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = n \cdot 1/n - 1.$$

Consequently, the average number of people who receive the correct hat is exactly 1.

Linearity of Expectation

Birthday Paradox: There are n people in a room, and m days in a year. What is the expected number of pairs of people who share the same birthday?

Let $X_{i,j}$ be a random variable that is 1 if person i and person j have the same birthday and 0 otherwise. Let X be number of pairs with same birthday. Then

$$X = \sum_{1 \le i < j \le n} X_{i,j}$$

By linearity of expectation

$$E(X) = \sum_{1 \le i < j \le n} E(X_{i,j})$$

Since $E(X_{i,j}) = 1/m$, we have that

$$E(X) = \frac{\binom{n}{2}}{m}$$

Quicksort

Divide: Pick an index q in the array. Partition the array into two subarrays, A1 and A2, where all elements in A1 are less than A[q], and all elements in A2 are greater than A[q]

Conquer: Recursively sort A1 and A2

Example

Quicksort the array [2, 6, 4, 1, 5, 3]

Quicksort

```
//PRE: A is the array to be sorted, p>=1;
//     r is <= the size of A
//POST: A[p..r] is in sorted order
Quicksort (A,p,r){
  if (p<r){
    q = Partition (A,p,r);
    Quicksort (A,p,q-1);
    Quicksort (A,q+1,r);
}
```

Quicksort

```
Partition (A,p,r){
  x = A[r];
  i = p-1;
  for (j=p;j<=r-1;j++){
    if (A[j]<=x){
      i++;
      exchange A[i] and A[j];
  }}
  exchange A[i+1] and A[r];
  return i+1;
}
```

Random Quicksort

Use the following R-Partition function

Protects against bad inputs

```
R-Partition (A,p,r){
  i = Random(p,r);
  exchange A[r] and A[i];
  return Partition(A,p,r);
}
```

Analysis

- Quicksort is a **randomized** algorithm
- The runtime is a **random** variable
- Want to compute the expected value

Analysis

- Let X be the number of comparisons made by Quicksort
- Approach: Will write X as the sum of a bunch of simpler random variables
- Then compute E(X) using linearity of expectation

Notation

- Let z_i be the the i-th smallest element in the array
- Let $Z_{i,j} = \{z_i\ z_{i+1}\ , \ldots, z_j\}$
- $X_{i,j} = 1$ if z_i and z_j are compared, 0 otherwise
- Then $$X = \sum_{1 \le i < j \le n} X_{i,j}$$

Analysis

- Q1: So what is $E(X_{i,j})$?
- A1: It is $P(z_i$ is compared to $z_j)$
- Q2: What is $P(z_i$ is compared to $z_j)$?
- A2: It is:

  P(either z_i or z_j are the first elems in $Z_{i,j}$ chosen as pivots)

- Why?
  - If no element in $Z_{i,j}$ has been chosen yet, no two elements in $Z_{i,j}$ have yet been compared, and all of $Z_{i,j}$ is in same list
  - If some element in $Z_{i,j}$ other than z_i or z_j is chosen first, z_i and z_j will be split into separate lists (and hence will never be compared)

- Q: What is

$$P(\text{either } z_i \text{ or } z_j \text{ are first elems in } Z_{i,j} \text{ chosen as pivots})$$

- A: $P(z_i \text{ chosen as first elem in } Z_{i,j}) +$
  $P(z_j \text{ chosen as first elem in } Z_{i,j})$
- Further note that number of elems in $Z_{i,j}$ is $j - i + 1$, so

$$P(z_i \text{ chosen as first elem in } Z_{i,j}) = \frac{1}{j-i+1}$$

and

$$P(z_j \text{ chosen as first elem in } Z_{i,j}) = \frac{1}{j-i+1}$$

- Hence

$$P(z_i \text{ or } z_j \text{ are first elems in } Z_{i,j} \text{ chosen as pivots}) = \frac{2}{j-i+1}$$

Putting it all Together

$$
\begin{aligned}
E(X) &= E\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}\right) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(X_{i,j}) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \\
&< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \\
&= \sum_{i=1}^{n-1} O(\log n) \\
&= O(n \log n)
\end{aligned}
$$

Putting it all Together

- Q: Why is $\sum_{k=1}^{n} \frac{2}{k} = O(\log n)$?
- A:

$$\begin{aligned} \sum_{k=1}^{n} \frac{2}{k} &= 2 \sum_{k=1}^{n} 1/k \\ &\leq 2(\ln n + 1) \end{aligned}$$

Where the last step holds by an integral bound (shown in-class)