Discrete Probability

Chapter 7

With Question/Answer Animations

Chapter Summary

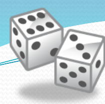
- Introduction to Discrete Probability
- Probability Theory
- Bayes' Theorem
- Expected Value and Variance

An Introduction to Discrete Probability

Section 7.1

Section Summary

- Finite Probability
- Probabilities of Complements and Unions of Events
- Probabilistic Reasoning



Probability of an Event

Pierre-Simon Laplace (1749-1827)

We first study Pierre-Simon Laplace's classical theory of probability, which he introduced in the 18th century, when he analyzed games of chance.

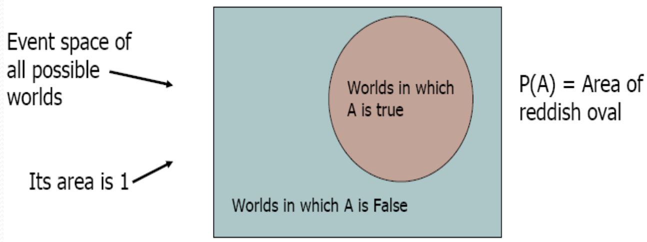
- We first define these key terms:
  - An *experiment* is a procedure that yields one of a given set of possible outcomes.
  - The *sample space* of the experiment is the set of possible outcomes.
  - An *event* is a subset of the sample space.
- Here is how Laplace defined the probability of an event:

  Definition: If S is a finite sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is $p(E) = |E|/|S|$.
- For every event E, we have $0 \leq p(E) \leq 1$. This follows directly from the definition because $0 \leq p(E) = |E|/|S| \leq |S|/|S| \leq 1$, since $0 \leq |E| \leq |S|$.

Probabilities

- We write P(A) as "the fraction of possible worlds in which A is true"
- We could at this point spend 2 hours on the philosophy of this.
- But we won't.

Visualizing A



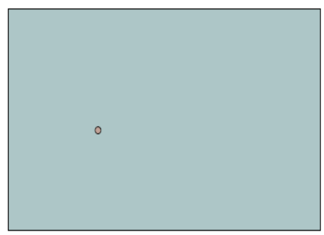
The Axioms of Probability

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

Where do these axioms come from? Were they "discovered"? Answers coming up later.

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)



The area of A can't get any smaller than 0

And a zero area would mean no world could ever have A true

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

The area of A can't get any bigger than 1

And an area of 1 would mean all worlds will have A true

Applying Laplace's Definition

Example: An urn contains four blue balls and five red balls. What is the probability that a ball chosen from the urn is blue?

Solution: The probability that the ball is chosen is 4/9 since there are nine possible outcomes, and four of these produce a blue ball.

Example: What is the probability that when two dice are rolled, the sum of the numbers on the two dice is 7?

Solution: By the product rule there are $6^2 = 36$ possible outcomes. Six of these sum to 7. Hence, the probability of obtaining a 7 is $6/36 = 1/6$.

Applying Laplace's Definition

Example: In a lottery, a player wins a large prize when they pick four digits that match, in correct order, four digits selected by a random mechanical process. What is the probability that a player wins the prize?

Solution: By the product rule there are 10,000 ways to pick four digits.

- Since there is only 1 way to pick the correct digits, the probability of winning the large prize is $1/10{,}000 = 0.0001$.

A smaller prize is won if only three digits are matched. What is the probability that a player wins the small prize?

Solution: If exactly three digits are matched, one of the four digits must be incorrect and the other three digits must be correct. For the digit that is incorrect, there are 9 possible choices. Hence, by the sum rule, there is a total of 36 possible ways to choose four digits that match exactly three of the winning four digits. The probability of winning the small price is $36/10{,}000 = 9/2500 = 0.0036$.

Applying Laplace's Definition

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 $1/\ 3{,}838{,}380 \approx 0.00000026$.

Can you work out the probability of winning the lottery with the biggest prize where you live?

Applying Laplace's Definition

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 $50 \cdot 49 \cdot 47 \cdot 46 = 254{,}251{,}200$ ways to choose the five balls.
b) *Sampling with replacement*: The probability is $1/50^5 = 1/312{,}500{,}000$ since $50^5 = 312{,}500{,}000$.

The Probability of Complements and Unions of Events

Theorem 1: Let E be an event in sample space S. The probability of the event $\overline{E} = S - E$, the complementary event of E, is given by

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

Proof: Using the fact that $|\overline{E}| = |S| - |E|$,

$$p(\overline{E}) = \frac{|S|-|E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E). \quad ◄$$

The Probability of Complements and Unions of Events

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,

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

The Probability of Complements and Unions of Events

Theorem 2: Let E_1 and E_2 be events in the sample space S. Then

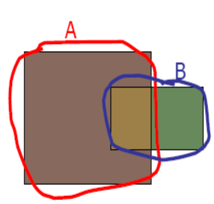
$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Proof: Given the inclusion-exclusion formula from Section 2.2, $|A \cup B| = |A| + |B| - |A \cap B|$, it follows that

$$\begin{aligned} p(E_1 \cup E_2) &= \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} \\ &= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} \\ &= p(E_1) + p(E_2) - p(E_1 \cap E_2). \end{aligned}$$ ◄

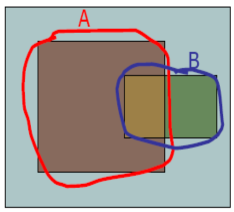
Interpreting the axioms

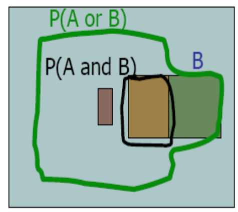
- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)



Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)





Simple addition and subtraction

The Probability of Complements and Unions of Events

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:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$
$$= 50/100 + 20/100 - 10/100 = 3/5.$$

Probability Theory

Section 7.2

Section Summary

- Assigning Probabilities
- Probabilities of Complements and Unions of Events
- Conditional Probability
- Independence
- Bernoulli Trials and the Binomial Distribution
- Random Variables
- The Birthday Problem
- Monte Carlo Algorithms
- The Probabilistic Method

Assigning Probabilities

Laplace's definition from the previous section, assumes that all outcomes are equally likely. Now we introduce a more general definition of probabilities that avoids this restriction.

- 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*.

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: 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. (Note that we could have used Laplace's definition here.)

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

Probability of an Event

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

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

- Note that now no assumption is being made about the distribution.

Example

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 $E = \{1,3,5\}$. Since $p(1) = p(2) = p(4) = p(5) = p(6)$ and $p(3) = 2p(1)$. Therefore, $5p(1) + p(3) = 5p(1) + 2p(1) = 1$. Thus, $p(1) = 1/7$ and $p(3) = 2/7$.

Hence,

$p(E) = p(1) + p(3) + p(5) = 1/7 + 2/7 + 1/7 = 4/7.$

Probabilities of Complements and Unions of Events

- Complements: $p(\overline{E}) = 1 - p(E)$ still holds. Since each outcome is in either E or $\overline{E}$, but not both,

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

- Unions: $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$ also still holds under the new definition.

Combinations of Events

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

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

see Exercises 36 and 37 for the proof

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