4190.101
Discrete Mathematics

Chapter 7 Discrete Probability

Gunhee Kim

[Based on S. Srinivasa Rao's Lecture Notes]

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 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 (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 since $0 \leq p(E) = |E|/|S| \leq |S|/|S| \leq 1$, since $0 \leq |E| \leq |S|$.

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 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 10^4 = 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 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 $1/\ 3{,}838{,}380 \approx 0.00000026.$

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 ·49 .48 ·47 ·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 $\bar{E} = S - E$, the complementary event of *E*, is given by

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

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

$$p(\bar{E}) = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E)$$

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 $\bar{E}$ is the event that all of the bits are 1s. The size of the sample space S is 2^{10}. Hence,

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

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

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

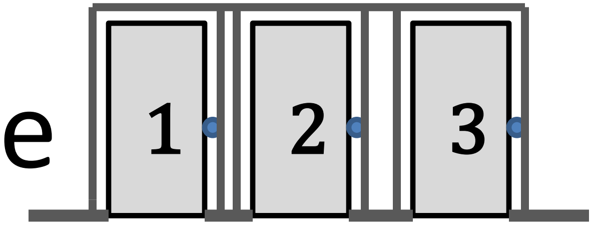
The Probability of Complements and Unions of Events

- **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:

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

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

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?
  - This is a notoriously confusing problem that has been the subject of much discussion.
- **Solution**: You should switch. The probability that your initial pick is correct is 1/3. This is the same whether or not you switch doors. But since the game show host always opens a door that does not have the prize, if you switch the probability of winning will be 2/3, because you win if your initial pick was not the correct door and the probability your initial pick was wrong is 2/3.

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 (*not currently included in the overheads*)

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:
  a) $0 \leq p(s) \leq 1$ for each $s \in S$
  b) $\sum_{s \in S} p(s) = 1$
- 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 $E = \{1,3,5\}$. We have $p(3) = 2/7$ and

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

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

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

Probabilities of Complements and Unions of Events

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

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

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

- Unions: 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* ∩ *F* = {0000, 0001, 0010, 0011, 0100}, *p*(*E*∩*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**: Suppose that all families of interest have two children. 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* ∩ *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

  $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and $m \geq 2$.

Bernoulli Trials

James Bernoulli
$(1854 - 1705)$

- **Definition**: Suppose an experiment can have only two possible outcomes, *e.g.*, the flipping of a coin or the random generation of a bit.
  - Each performance of the experiment is called a *Bernoulli trial*.
  - One outcome is called a *success* and the other a *failure*.
  - If *p* is the probability of success and *q* the probability of failure, then *p* + *q* = 1.
- Many problems involve determining the probability of *k* successes when an experiment consists of *n* mutually independent Bernoulli trials.

Bernoulli Trials

- **Example**: A coin is biased so that the probability of heads is 2/3. What is the probability that exactly four heads occur when the coin is flipped seven times?
- **Solution**: There are $2^7 = 128$ possible outcomes. The number of ways four of the seven flips can be heads is $C(7,4)$. The probability of each of the outcomes is $(2/3)^4(1/3)^3$ since the seven flips are independent. Hence, the probability that exactly four heads occur is

$$C(7,4)\ (2/3)^4(1/3)^3 = (35 \cdot 16)/\ 2^7 = 560/\ 2187.$$

Probability of *k* Successes in *n* Independent Bernoulli Trials

- **Theorem 2**: The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is $C(n,k)p^kq^{n-k}$.
- **Proof**: The outcome of n Bernoulli trials is an n-tuple $(t_1,t_2,\ldots,t_n)$, where each is t_i either S (success) or F (failure). The probability of each outcome of n trials consisting of k successes and n - k failures (in any order) is p^kq^{n-k}. Because there are $C(n,k)$ n-tuples of Ss and Fs that contain exactly k Ss, the probability of k successes is $C(n,k)p^kq^{n-k}$. ◀
- We denote by $b(k{:}n,p)$ the probability of k successes in n independent Bernoulli trials with p the probability of success. Viewed as a function of k, $b(k{:}n,p)$ is the *binomial distribution*. By Theorem 2,

$$b(k{:}n,p) = C(n,k)p^kq^{n-k}.$$

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.
  - A random variable is a function. It is not a variable, and it is not random!
- **Example**: Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads when t is the outcome. Then $X(t)$ takes on the following values:
  - $X(HHH) = 3$, $X(TTT) = 0$, $X(HHT) = X(HTH) = X(THH) = 2$, $X(TTH) = X(THT) = X(HTT) = 1$.

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**: (from the same experiment in previous slide) 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 $p(X = 0) = 1/8$.

The Famous Birthday Problem

- What is the minimum number of people who need to be in a room so that the probability that at least two of them have the same birthday is greater than 1/2?
- **Solution**: We assume that all birthdays are equally likely and that there are 366 days in the year. First, we find the probability p_n that n people have different birthdays.
  - The probability that the birthday of the second person is different from that of the first is 365/366.
  - The probability that the birthday of the third person is different from the other two, when these have two different birthdays, is 364/366.
  - In general, the probability that the j-th person has a birthday different from the birthdays of those already in the room, assuming that these people all have different birthdays, is $(366 - (j - 1))/366 = (367 - j)/366$.
  - Hence, $p_n = (365/366)(364/366)\cdots (367 - n)/366$.
  - Therefore , $1- p_n = 1-(365/366)(364/366)\cdots (367 - n)/366$.
- Checking various values for n with computation help tells us that for $n = 22$, $1- p_n \approx 0.457$, and for $n = 23$, $1- p_n \approx 0.506$. Consequently, a minimum number of 23 people are needed so that that the probability that at least two of them have the same birthday is greater than 1/2.

Monte Carlo Algorithms

- Algorithms that make random choices at one or more steps are called *probabilistic algorithms*.
- *Monte Carlo algorithms* are probabilistic algorithms used to answer decision problems, which are problems that either have “true” or “false” as their answer.
  - A Monte Carlo algorithm consists of a sequence of tests. For each test the algorithm responds “true” or “unknown.”
  - If the response is “true,” the algorithm terminates with the answer is “true.”
  - After running a specified sequence of tests where every step yields “unknown”, the algorithm outputs “false.”
  - The idea is that the probability of the algorithm incorrectly outputting “false” should be very small as long as a sufficient number of tests are performed.

Probabilistic Primality Testing

- Probabilistic primality testing (*see Example* 16 *in text*) is an example of a Monte Carlo algorithm, which is used to find large primes to generate the encryption keys for RSA cryptography (*as discussed in Chapter* 4).
  - An integer n greater than 1 can be shown to be composite (i.e., not prime) if it fails a particular test (Miller's test), using a random integer b with $1 < b < n$ as the base. The probability that a composite integer passes n Miller's test is for a random b, is less that ¼.
  - So failing the test, is the "true" response in a Monte Carlo algorithm, and passing the test is "unknown."
  - If the test is performed k times (choosing a random integer b each time) and the number n passes Miller's test at every iteration, then the probability that it is composite is less than $(1/4)^k$. So for a sufficiently, large k, the probability that n is composite even though it has passed all k iterations of Miller's test is small. For example, with 10 iterations, the probability that n is composite is less than 1 in 1,000,000.