Discrete Probability

Dit-Yan Yeung

Department of Computer Science and Engineering
Hong Kong University of Science and Technology

COMP 2711: Discrete Mathematical Tools for Computer Science

Contents

Introduction

In computer science, probability theory plays an important role in many applications.

One of them is **hashing**.

Another important application is in the study of the complexity of algorithms. In particular, ideas and techniques from probability theory are used to determine the **average-case complexity of algorithms**.

Probabilistic algorithms can be used to solve many problems that cannot be easily or practically solved by deterministic algorithms.

Probability of an Event

Definition

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 all possible outcomes. An **event** is a subset of the sample space.

Definition

If S is a finite sample space of equally likely outcomes, and E is an event, i.e., a subset of S, then the **probability** of E is

$$p(E) = \frac{|E|}{|S|}.$$

Examples

Example 1

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

Example 2

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

Examples

Example 3

In a lottery, players win a large prize when they pick four digits that match, in the correct order, four digits selected by a random mechanical process. A smaller prize is won if only three digits are matched. What is the probability that a player wins the large prize? What is the probability that a player wins the small prize?

Example 4

In a lottery, a player wins an enormous prize if he correctly chooses a set of six numbers out of the first 40 positive integers. What is the probability that a player chooses the correct six numbers?

Examples

Example 5

Find the probability that a hand of five cards in poker contains four cards of one kind.

Example 6

What is the probability that a poker hand contains a full house, i.e., three of one kind and two of another kind?

Example 7

What is the probability that the numbers 11, 4, 17, 39, and 23 are drawn in that order from a bin containing 50 balls labeled with the numbers $1, 2, \ldots, 50$ if (a) the ball selected is not returned to the bin before the next ball is selected, and (b) the ball selected is returned to the bin before the next ball is selected?

Complement

We can use counting techniques to find the probability of events derived from other events.

Theorem 1.1

Let E be an event in a finite sample space S. The probability of the event $\overline{E}$, the complementary event of E, is given by $p(\overline{E}) = 1 - p(E)$.

Proof.

Note that $|\overline{E}| = |S| - |E|$. Hence,

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

□

Example

There is an alternative strategy for finding the probability of an event when a direct approach does not work well.

Instead of determining the probability of the event, the probability of its complement can be found.

Example 8

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

Union

Theorem 1.2

Let E_1 and E_2 be events in a sample space S. Then

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

Proof

Proof.

Recall that

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|.$$

Hence,

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

□

Example

Example 9

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?

Probability Distribution

Definition

Let S be the sample space of an experiment with a finite or countably infinite number of outcomes. A **probability distribution** on S is characterized by a **probability mass function** p from S such that

(a) $0 \leq p(s) \leq 1$ for each $s \in S$

(b) $\sum_{s \in S} p(s) = 1$,

where $p(s)$ is the **probability of an outcome** s.

To model an experiment, the probability $p(s)$ assigned to an outcome s should equal the limit of the number of times s occurs divided by the number of times the experiment is performed, as this number grows without bound.

Example

Example 10

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

Uniform Distribution

Definition

Let S be a set with n elements. The **uniform distribution** assigns the probability $1/n$ to each element of S.

The experiment of selecting an element from a sample space with a uniform distribution is called selecting an element of S **at random**.

Probability of an Event

Definition

The **probability of an event** E is the sum of the probabilities of the outcomes in E, i.e.,

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

Example 11

Suppose a dice is biased such 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 dice?

Remark

Remark

For a finite sample space with n equally likely outcomes, the probability of an event E that contains m outcomes is

$$p(E) = \sum_{s \in E} p(s) = \sum_{s \in E} \frac{1}{n} = \frac{m}{n} = \frac{|E|}{|S|}.$$

Complement

Theorem 2.1

Let E be an event in a finite or countably infinite sample space S. The probability of the complementary event $\overline{E}$ is given by $p(\overline{E}) = 1 - p(E)$.

Proof.

Because $E \cup \overline{E} = S$ and $E \cap \overline{E} = \emptyset$,

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

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

Inclusion-Exclusion Principle for Probability

Theorem 2.2 (Inclusion-exclusion principle for probability)

Let $E_1, E_2, \ldots, E_n$ be events in a sample space S. Then

$$p\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{1 \le i \le n} p(E_i) - \sum_{1 \le i < j \le n} p(E_i \cap E_j) + \sum_{1 \le i < j < k \le n} p(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n+1} p(E_1 \cap E_2 \cap \cdots \cap E_n).$$

Proof.

The proof is similar to that for counting. □

Corollary

Corollary 2.3

Let $E_1, E_2, \ldots, E_n$ be pairwise disjoint events in a sample space S. Then

$$p\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{1 \le i \le n} p(E_i).$$

Proof.

Since the events are pairwise disjoint, i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j$, all terms except the first one on the right-hand side of the above theorem are equal to 0. □

Conditional Probability

Definition

Let E and F be events with $p(F) > 0$. The **conditional probability** of E given F, denoted by $p(E \mid F)$, is defined as

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

Remark

It is like taking F as the sample space. For an outcome from E to occur, it must also belong to $E \cap F$.

Remark

If $p(F) = 0$, F gives us no new information about our situation. Thus we define $p(E \mid F) = p(E)$ when $p(F) = 0$.

Examples

Example 12

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

Example 13

What is the conditional probability that a family with two children has two boys, given 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. Note that BG represents a family with an older boy and a younger girl while GB represents a family with an older girl and a younger boy.

Example

Example 14

If a student knows 80% of the material in a course, what is the probability that she answers a question correctly on a well-balanced true-false test assuming that she randomly guesses on any question for which she does not know the answer?

Independence

Definition

Two events E and F are **independent** if and only if $p(E \cap F) = p(E)p(F)$ or, equivalently, $p(E \mid F) = p(E)$.

Example 15

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?

Examples

Example 16

Assume that each of the four ways a family can have two children 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?

Example 17

Are the events E, that a family with three children has children of both sexes, and F, that this family has at most one boy, independent? Assume that the eight ways a family can have three children are equally likely.

Bernoulli Trial

Definition

Each performance of an experiment with two possible outcomes is called a **Bernoulli trial**. The two possible outcomes are generally referred to as **success** and **failure**.

Many problems can be solved by determining the probability of k successes when an experiment consists of n mutually independent Bernoulli trials.

Example 18

A coin is biased so that the probability of a head is $2/3$. What is the probability that exactly four heads come up when the coin is flipped seven times, assuming that the flips are independent?

Theorem

Theorem 2.4

The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is

$$\binom{n}{k} p^k q^{n-k}.$$

Proof.

When n Bernoulli trials are carried out, the outcome can be represented as an n-tuple $(t_1, t_2, \ldots, t_n)$, where $t_i = S$ (for success) or $t_i = F$ (for failure) for $i = 1, 2, \ldots, n$. Because the n trials are independent, the probability of each outcome of n trials consisting of k successes and $n - k$ failures in any order is $p^k q^{n-k}$. Because there are $C(n, k)$ n-tuples of S's and F's that contain k S's, the probability of k successes is $\binom{n}{k} p^k q^{n-k}$. □

Binomial Distribution

Definition

The probability of k successes in n independent Bernoulli trials with probability of success p and probability of failure $1 - p$, denoted by $b(k; n, p)$, is

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}.$$

When considered as a function of k, it is called the **binomial distribution**.

Examples

Example 19

A student takes a 10-question test. Suppose that the student who knows 80% of the course material has probability 0.8 of success on any question, independent of her performance on any other question. What is the probability that she gets the correct answers for at least eight questions?

Example 20

Suppose the probability that a 0 bit is generated is 0.9, and the probability that a 1 bit is generated is 0.1, and that the bits are generated independently. What is the probability that exactly eight 0 bits are generated when 10 bits are generated?

Remark

Remark

The sum of the probabilities that there are k successes when n independent Bernoulli trials are carried out, for $k = 0, 1, 2, \ldots, n$, is equal to

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

By the binomial theorem,

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = (p+1-p)^n = 1.$$

Random Variable

Definition

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

Remark

A random variable is a function, not a variable.

Example

Example 21

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

$$\begin{aligned} X(HHH) &= 3 \\ X(HHT) = X(HTH) = X(THH) &= 2 \\ X(TTH) = X(THT) = X(HTT) &= 1 \\ X(TTT) &= 0. \end{aligned}$$

Distribution of a Random Variable

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. A distribution is usually described by specifying $p(X = r)$ for each $r \in X(S)$.

Examples

Example 22

Because each of the eight possible outcomes when three coins are flipped has probability $1/8$, the distribution of the random variable $X(t)$ in the example above is given by $P(X = 3) = 1/8$, $P(X = 2) = 3/8$, $P(X = 1) = 3/8$, $P(X = 0) = 1/8$.

Example 23

Let X be the sum of the numbers that appear when a pair of dice is rolled. What are the values of this random variable for the 36 possible outcomes (i, j), where i and j are the numbers that appear on the first dice and the second dice, respectively, when these two dice are rolled?

Birthday Problem

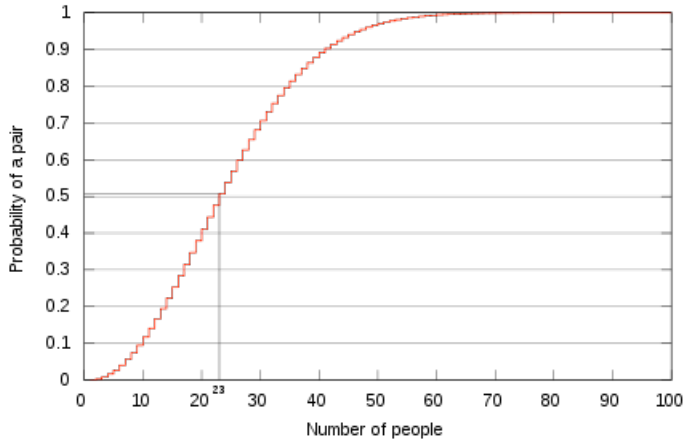
Example 24

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 1, assuming that there are 366 days in a year?

Example 25 (Birthday problem or birthday paradox)

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 $1/2$, assuming that there are 366 days in a year, the birthdays of the people in the room are independent, and each birthday is equally likely?

Birthday Problem (cont'd)



Hashing

A **hash function** $h(k)$ is a mapping of the keys (of the records that are to be stored in a database) to storage locations, where the number of possible keys is much larger than the number of available storage locations in the hash table.

A good hash function should yield few **collisions**, which are mappings of two different keys to the same memory location, when relatively few of the records are in play in a given application.

Example

Example 26

What is the probability that there are no collisions in a hash table with m available storage locations, i.e., the probability that no two keys are mapped to the same location? We assume that the keys of the records selected have an equal probability to be any of the elements in the key space, these keys are selected independently, and the hash function distributes the keys uniformly to the m storage locations.

Bayes' Theorem

There are times when we want to assess the probability that a particular event occurs on the basis of partial evidence.

Theorem 3.1 (Bayes' theorem)

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

$$p(F \mid E) = \frac{p(E \mid F)\, p(F)}{p(E \mid F)\, p(F) + p(E \mid \overline{F})\, p(\overline{F})}.$$

Proof

Proof of Bayes' theorem.

The definition of conditional probability tells us that

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

Therefore,

$$p(F \mid E)\, p(E) = p(E \mid F)\, p(F).$$

Dividing both sides by $p(E)$, we find that

$$p(F \mid E) = \frac{p(E \mid F)\, p(F)}{p(E)}.$$

Proof (cont'd)

Proof (cont'd).

To complete the proof, we need to show that

$$p(E) = p(E \mid F)\, p(F) + p(E \mid \overline{F})\, p(\overline{F}).$$

First, note that

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

Furthermore, we note that $E \cap F$ and $E \cap \overline{F}$ are disjoint. Consequently, $p(E) = p(E \cap F) + p(E \cap \overline{F})$. We have already shown that $p(E \cap F) = p(E \mid F)\, p(F)$. Moreover, we have $p(E \mid \overline{F}) = p(E \cap \overline{F})/p(\overline{F})$ and hence $p(E \cap \overline{F}) = p(E \mid \overline{F})\, p(\overline{F})$. It thus follows that

$$p(E) = p(E \cap F) + p(E \cap \overline{F}) = p(E \mid F)\, p(F) + p(E \mid \overline{F})\, p(\overline{F}).$$

Example

Example 27

We have two boxes. The first box contains two green balls and seven red balls; the second one contains four green balls and three red balls. Bob selects a ball by first choosing one of the two boxes at random. He then selects one of the balls in this box at random. If Bob has selected a red ball, what is the probability that he selected a ball from the first box?

Example

Example 28

Suppose one person in 100,000 has a particular rare disease for which there is a fairly accurate diagnostic test. This test is correct 99% of the time when given to someone with the disease; it is correct 99.5% of the time when given to someone who does not have the disease. Given this information, can we find

(a) the probability that someone who tests positive for the disease has the disease?

(b) the probability that someone who tests negative for the disease does not have the disease?

Should someone who tests positive be very concerned that he or she has the disease?

Generalized Bayes' Theorem

Theorem 3.2 (Generalized Bayes' theorem)

Suppose E is an event from a sample space S and $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$. Assume that $p(E) \neq 0$ and $p(F_i) \neq 0$ for $i = 1, 2, \ldots, n$. Then

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

Bayesian Spam Filter

A **Bayesian spam filter**, based on Bayes' theorem, uses information about previously seen e-mail messages to guess whether an incoming e-mail message is spam.

Example 29

Suppose we have found that the word 'Rolex' occurs in 250 of 2,000 messages known to be spam and in 5 of 1,000 messages known not to be spam. 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 rejecting a message as spam is 0.9, will we reject this message?

Example

Example 30

Suppose we train a Bayesian spam filter on a set of 2,000 spam messages and 1,000 messages that are not spam. The word 'stock' appears in 400 spam messages and 60 messages that are not spam, and the word 'undervalued' appears in 200 spam messages and 25 messages that are not spam. Estimate the probability that an incoming message containing both the words 'stock' and 'undervalued' is spam, assuming that we have no prior knowledge about whether it is spam. Will we reject the message as spam when we set the threshold at 0.9?

Remark

In general, the more words we use to estimate the probability that an incoming mail message is spam, the better is our chance that we correctly determine whether it is spam.

Expected Value

Definition

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

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

Remark

When there are infinitely many elements in the sample space, the expectation is defined only when the infinite series in the definition is absolutely convergent. In particular, the expectation of a random variable on an infinite sample space is finite if it exists.

Examples

Example 31

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

Example 32

A fair coin is flipped three times. Let S be the sample space of the eight possible outcomes and let X be the random variable that assigns to an outcome the number of heads in this outcome. What is the expected value of X?

Expected Value (cont'd)

When an experiment has a large number of outcomes, it may be inconvenient to compute the expected value of a random variable directly from its definition. Instead, we can find the expected value of a random variable by grouping together all outcomes assigned the same value by the random variable.

Theorem 4.1

If X is a random variable and $p(X = r)$ is the probability that $X = r$, so that $p(X = r) = \sum_{s \in S, X(s)=r} p(s)$, then

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

Proof

Proof.

Suppose X is a random variable with range $X(S)$. Let $p(X = r)$ be the probability that the random variable X takes the value r. Consequently, $p(X = r)$ is the sum of the probabilities of the outcomes s such that $X(s) = r$. It follows that

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

□

Example 33

What is the expected value of the sum of the numbers that appear when a pair of fair dice is rolled?

Some Results

Proposition 4.1

If X is a random variable that always takes on a constant value c, then $E(X) = c$.

Proof.

$E(X) = c \cdot p(X = c) = c \cdot 1 = c.$ □

Corollary 4.2

Let X be a random variable on a sample space. Then $E\big(E(X)\big) = E(X)$.

Proof.

When we think of $E(X)$ as a random variable, it has a constant value traditionally denoted by μ. By Proposition 4.1, we have that $E\big(E(X)\big) = E(\mu) = \mu = E(X)$. □

Average-Case Computational Complexity

Computing the average-case computational complexity of an algorithm can be interpreted as computing the expected value of a random variable.

Definition

Let the sample space of an experiment be the set of possible inputs a_i, $i = 1, 2, \ldots, n$, and let X be the random variable that assigns to a_i the number of operations used by the algorithm when given a_i as input. Based on our knowledge of the input, we assign a probability $p(a_i)$ to each possible input value a_i. Then, the **average-case computational complexity** of the algorithm is

$$E(X) = \sum_{i=1}^{n} p(a_i) X(a_i),$$

which is the expected value of X.

Linearity of Expectation

Theorem 4.3

If X_i, $i = 1, 2, \ldots, n$, are random variables on S, and if a and b are real numbers, then

(i) $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$

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

Proof

Proof.

The first result follows directly from the definition of expected value, because

$$\begin{aligned} & E(X_1 + X_2 + \cdots + X_n) \\ &= \sum_{s \in S} p(s)\Big[X_1(s) + X_2(s) + \cdots + X_n(s)\Big] \\ &= \sum_{s \in S} p(s)X_1(s) + \sum_{s \in S} p(s)X_2(s) + \cdots + \sum_{s \in S} p(s)X_n(s) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n). \end{aligned}$$

Proof (cont'd)

Proof (cont'd).

For the second result, note that

$$\begin{aligned} E(aX + b) &= \sum_{s \in S} p(s)\left[aX(s) + b\right] \\ &= a\sum_{s \in S} p(s)X(s) + b\sum_{s \in S} p(s) \\ &= a\,E(X) + b, \end{aligned}$$

because $\sum_{s \in S} p(s) = 1$. □

Example 34

Using Theorem 4.3, find the expected value of the sum of the numbers that appear when a pair of fair dice is rolled.

Expected Number of Successes

Theorem 4.4

The expected number of successes when n independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Proof.

Let X_i be the random variable with $X_i((t_1, t_2, \ldots, t_n)) = 1$ if t_i is a success and $X_i((t_1, t_2, \ldots, t_n)) = 0$ if t_i is a failure. The expected value of X_i is $E(X_i) = 1 \cdot p + 0 \cdot (1 - p) = p$ for $i = 1, 2, \ldots, n$. Let $X = X_1 + X_2 + \cdots + X_n$, so that X counts the number of successes when these n Bernoulli trials are performed. Thus,
$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$. □

Example

Example 35

A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the expected number of hats that are returned correctly?

Indicator Random Variable

Definition

A random variable that is equal to 1 if a certain event occurs and 0 otherwise is called an **indicator random variable**.

Remark

An indicator random variable X has the nice property that

$$E(X) = p(X = 1) = p(\text{the event occurs}).$$

Geometric Distribution

We now look at a random variable with infinitely many possible outcomes.

Example 36

Suppose the probability that a coin comes up with a tail is p. This coin is flipped repeatedly until it comes up with a tail. What is the expected number of flips until this coin comes up with a tail?

Remark

If the coin is fair, we have $p = 1/2$ and so the expected number of flips until a tail comes up is $1/(1/2) = 2$.

Geometric Distribution (cont'd)

Definition

A random variable X has a **geometric distribution** with parameter p if $p(X = k) = (1 - p)^{k-1}p$ for $k = 1, 2, 3, \ldots$.

Theorem 4.5

If the random variable X has the geometric distribution with parameter p, then $E(X) = 1/p$.

Independent Random Variables

Definition

The random variables X and Y on a sample space S are **independent** if

$$p(X(s) = r_1 \text{ and } Y(s) = r_2) = p(X(s) = r_1) \cdot p(Y(s) = r_2),$$

for all real numbers r_1 and r_2.

Examples

Example 37

A pair of fair dice is rolled. Let X_1 and X_2 be the random variables for the numbers appearing on the first and second dice, respectively. Are X_1 and X_2 independent?

Example 38

For the problem above, show that the random variables X_1 and $X = X_1 + X_2$ are not independent.

Theorem

Theorem 4.6

If X and Y are independent random variables on a space S, then $E(XY) = E(X)E(Y)$.

Proof

Proof.

From the definition of expected value and because X and Y are independent random variables, it follows that

$$\begin{aligned}
E(XY) &= \sum_{s \in S} X(s)Y(s)\,p(s) \\
&= \sum_{r_1 \in X(S), r_2 \in Y(S)} r_1 r_2 \cdot p\big(X(s) = r_1 \text{ and } Y(s) = r_2\big) \\
&= \sum_{r_1 \in X(S), r_2 \in Y(S)} r_1 r_2 \cdot p\big(X(s) = r_1\big) \cdot p\big(Y(s) = r_2\big) \\
&= \left[\sum_{r_1 \in X(S)} r_1\, p\big(X(s) = r_1\big)\right] \cdot \left[\sum_{r_2 \in Y(S)} r_2\, p\big(Y(s) = r_2\big)\right] \\
&= E(X)E(Y).
\end{aligned}$$

□

Example

Example 39

Let X and Y be random variables that count the number of heads and the number of tails when a coin is flipped twice. Show that X and Y are not independent.

Variance and Standard Deviation

The expected value of a random variable tells us its average value, but nothing about how widely its values are distributed. The variance of a random variable helps us characterize how widely a random variable is distributed.

Definition

Let X be a random variable on a sample space S. The **variance** of X, denoted by $V(X)$, is

$$V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s).$$

The **standard deviation** of X, denoted by $\sigma(X)$, is defined to be $\sqrt{V(X)}$.

Theorem

Theorem 4.7

If X is a random variable on a sample space S, then $V(X) = E(X^2) - E(X)^2$.

Proof.

Note that

$$\begin{aligned} V(X) &= \sum_{s \in S} (X(s) - E(X))^2 p(s) \\ &= \sum_{s \in S} X^2(s)p(s) - 2E(X) \sum_{s \in S} X(s)p(s) + E(X)^2 \sum_{s \in S} p(s) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

□

Examples

Example 40

What is the variance of the random variable X with $X(t) = 1$ if a Bernoulli trial is a success and $X(t) = 0$ if it is a failure, where p is the probability of success?

Example 41

What is the variance of the random variable X, where X is the number that comes up when a dice is rolled?

Example 42

What is the variance of the random variable $X((i,j)) = 2i$, where i is the number appearing on the first dice and j is the number appearing on the second dice, when two dice are rolled?

Theorem

Theorem 4.8

If X and Y are two independent random variables on a sample space S, then $V(X+Y) = V(X) + V(Y)$. Furthermore, if $X_i, i = 1, 2, \ldots, n$, with n a positive integer, are pairwise independent random variables on S, then $V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$.

Proof

Proof.

We note that

$$\begin{aligned} V(X+Y) &= E\big((X+Y)^2\big) - E(X+Y)^2 \\ &= E(X^2 + 2XY + Y^2) - \big(E(X) + E(Y)\big)^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2. \end{aligned}$$

Because X and Y are independent, we have $E(XY) = E(X)E(Y)$. It follows that

$$V(X+Y) = \big(E(X^2) - E(X)^2\big) + \big(E(Y^2) - E(Y)^2\big) = V(X) + V(Y).$$

The proof of the case with n pairwise independent random variables can be constructed by generalizing the proof of the case for two random variables. □

Examples

Example 43

Find the variance and standard deviation of the random variable X whose value when two dice are rolled is $X((i,j)) = i + j$, where i is the number appearing on the first dice and j is the number appearing on the second dice.

Example 44

What is the variance of the number of successes when n independent Bernoulli trials are performed, where p is the probability of success on each trial?

Expected Number of Items per Location

Theorem 5.1

In hashing n items into a hash table with m locations, the expected number of items that hash to any one location is n/m.

Proof

Proof.

Without loss of generality, we choose a specific location, say location 1, and consider hashing items into it. We consider the process of hashing n items into the hash table as n independent trials, each with m possible outcomes corresponding to the m locations in the hash table. Let X_i be an indicator random variable that is 1 if, in the ith trial, the item hashes to location 1, and 0 otherwise. Note that the probability of hashing any one item into location 1 is $1/m$ because all m locations are equally likely. It follows that $E(X_i) = 1/m$. Let X be a random variable defined as $X = X_1 + X_2 + \cdots + X_n$. The expected number of items in location 1 is $E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = n/m$. There is nothing special about location 1 and so this expected value applies to any location. □

Expected Number of Empty Locations

Theorem 5.2

In hashing n items into a hash table with m locations, the expected number of empty locations is $m(1 - 1/m)^n$.

Proof

Proof.

The probability that location i is empty after hashing n items is $(1 - 1/m)^n$. Let X_i be an indicator random variable that is 1 if location i is empty after hashing n items and 0 otherwise. Thus $E(X_i) = (1 - 1/m)^n$. Let X be a random variable defined as $X = X_1 + X_2 + \cdots + X_m$. The expected number of empty locations after hashing n items is $E(X) = E(X_1 + X_2 + \cdots + X_m) = E(X_1) + E(X_2) + \cdots + E(X_m) = m(1 - 1/m)^n$. □

Remark

Remark

If $m = n$, i.e., we hash n items into a hash table with n locations, then the expected number of empty locations is $n(1 - 1/n)^n$. As n increases, we have

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e}$$

and hence the expected number of empty locations is n/e.

Expected Number of Collisions

Theorem 5.3

In hashing n items into a hash table with m locations, the expected number of collisions is $n - m + m(1 - 1/m)^n$.

Proof.

Let X_c, X_o and X_e be random variables for the number of collisions, the number of occupied locations, and the number of empty locations, respectively. We note that

$$X_c = n - X_o = n - (m - X_e) = n - m + X_e.$$

Thus the expected number of collisions is
$E(X_c) = n - m + E(X_e) = n - m + m(1 - 1/m)^n$. □

Expected Number of Items until all Locations have at least One Item

Theorem 5.4

The expected number of items needed to fill all locations of a hash table of size m is between $m \ln m + m/4$ and $m \ln m + m$.

Proof

Proof.

Let X_i be a random variable for the number of items added between the time when there are $i - 1$ full locations for the first time and i full locations for the first time.
We note that $X_1 = 1$ and hence $E(X_1) = 1$. To compute $E(X_2)$, we note that X_2 counts the number of independent trials until the first success (i.e., hashing an item to an unused location) with the probability of success of each trial equal to $(m - 1)/m$. Thus, $E(X_2) = m/(m - 1)$. In general, $E(X_i) = m/(m - i + 1)$.

Proof (cont'd)

Proof (cont'd).

Consequently, the expected number of items needed to fill all locations is

$$\begin{aligned} E(X_1 + X_2 + \cdots + X_m) &= E(X_1) + E(X_2) + \cdots + E(X_m) \\ &= \sum_{k=1}^{m} \frac{m}{m - k + 1} \\ &= m \sum_{k=1}^{m} \frac{1}{m - k + 1} = m \sum_{k=1}^{m} \frac{1}{k} = m H_m, \end{aligned}$$

where $H_m = \sum_{k=1}^{m} (1/k)$ is a **harmonic number**. It is known that

$$\ln m + \frac{1}{4} \leq H_m \leq \ln m + 1.$$

Therefore, $m \ln m + \frac{m}{4} \leq m H_m \leq m \ln m + m$. □

Remark

Remark

So, to fill every location in a hash table of size m, we need to hash roughly $m \ln m$ items. This problem is sometimes called the **coupon collector's problem**.