Soumil Nitin Shah

Python Developer

With Experience in Python; focused on using cutting-edge technologies to solve challenging problems: I turn ideas into projects, and projects into serial success. I Develop Python Based Cross Platform Desktop Application, Webpages , Software, REST API, Database

Website : - https://soumilshah.herokuapp.com

soushah@my.bridgeport.edu

+1(646)-204 -5957

Bridgeport, USA

https://www.linkedin.com/in/shah-soumil

https://github.com/soumilshah1995

https://www.youtube.com/channel/UC_eOodxvwS_H7x2uLQa-svw

Blog

EDUCATION

Master of Science Electrical engineering
Master of Science Computer Engineering

University of Bridgeport

Bachelor in Electronic Engineering

University of Mumbai *01/2014 – 11/2016*

WORK EXPERIENCE

Security Analyst Intern

Outsecure Inc

Shelton Connecticut

- Assist with vulnerability assessments and penetration testing for specific applications, services, networks and servers as required.
- Developed Python Shell Scripts for company that Automatically generates Report of all connected Device with their IP and Mac Address and open Ports. The Shell script can also find all vulnerable Devices such as Chrome cast and can execute Payload remotely.

Research & Teaching Assistant.

University of Bridgeport

Bridgeport June 2018- January 2019

- Developed a Proof of Concept of SMART Mirror (System Demo) using Raspberry Pi.
- Developed a device – 'Airsense' (System Demo) that maps pollutions and sends data to cloud server for further analysis
- Helped Dr.Abdel Shakour to conduct Labs and lecture for subject Internet-of-Things
- Developed new paradigm of monitoring student attendance using RFID based on the Internet of Thing (IoT) (System Demo)

Independent Researcher

Bridgeport June 2018- January 2019

University of Bridgeport

- Worked for a **NASA CT Space Grant Project** '*Balluino: High Altitude IOT Based Real Time Air Quality Management System Using Balloon/Drone' (System Demo)*

Embedded Hardware Developer Intern

Software Development Cell

India

- Developed a low-cost device that measures the pH level of soil in just 10% of the cost compared to Market Price.
- Developed an Embedded system that automated the existing light control using GSM, thereby reducing the human effort for rural areas.
- Designed a timer that can trigger switch control for Media preparation room Autoclave Controller in Tissue.

Projects

TensorFlow Projects

1. Machine Learning Model for Diabetes Prediction [Link]
2. Machine Learning Model on Wine Dataset [Link]
3. Machine Learning on Chest X-ray to Detect Pneumonia[Link]
4. Machine Learning for Housing Price Predictor [Link]
5. Machine learning model on IRIS Dataset [Link]

Library Developed in Python for Developers

1. This is smart library developed in python. The Library allows Developers to upload and retrieve data from cloud like Thingspeak, allows custom notification using SMS and Email using IFTTT [Link]
2. Python based Random Proxy [Link]

Python based Audio Book

1. Python based Audio book allows user to convert their Favorite Story book into a MP3 File. It allows conversion of text to MP3 File as well. Supports Language English and French.
2. This project we used Rest API and Text to speech Conversion and front end was developed in PyQt5 and Mongo DB was used as Database

Python Based Library Room Reservation system

1. This project was developed in Python using Twilio Phone call API. Over the Phone call user can check Booking, Reserve Room, Get notified about activity going on campus and automated Form Filling.[Link]

Python Certification

1. The Complete Python 3 Course: Go from Beginner to Advanced!
2. The Complete SQL Bootcamp
3. Python and Flask Bootcamp: Create Websites using Flask!
4. Learn Python Programming Masterclass
5. Introduction to Databases and SQL Querying
6. Deep Learning: An Introduction
7. C# Basics for Beginners: Learn C# Fundamentals by Coding

Publication

1. Teaching Internet-of-Things Using E-Learning Laboratory [Link]
2. Balluino: High Altitude Balloon-based Arduino Real Time Air Quality Monitoring System [Link]
3. IoT Based Smart Attendance System (SAS) Using RFID [Link]
4. Simulation of PM2.5 Particulate Matter Pollution in US East Coast Using SMAT-CE Software [Link]

Python Skills

Python Web Framework: HTML, CSS, JavaScript, jQuery, Bootstrap, Flask.
Python GUI Framework: Tkinter, PyQT5, Kivy, Electron JS (Eel).
Python Database Framework: Sqlite3, MySQL, SQL, Mongo DB
Python Data Science Framework: Pandas, Matplotlib, Seaborn
Python Web Scrapping Framework: Beautiful Soup, Selenium
Languages: C, C++, Embedded C, Verilog, C#

Honors

1. Best Academic Achievement Award [Link]
2. 3rd Prize at Hackathon Competition [Link]
3. Academic Awards 5000 USD

[CSED261] Discrete Mathematics for Computer Science

Lecture 10

Basic Number Theory

Suha Kwak

suha.kwak@postech.ac.kr

Dept. of Computer Science and Engineering

POSTECH

Divisibility and Modular Arithmetic

Division

- Definition
  - If a and b are integers with $a \neq 0$, then a divides b if there exists an integer c such that $b = ac$.
  - When a divides b, we say that a is a factor or divisor of b and that b is a multiple of a.

- Notation
  - The notation $a \mid b$ denotes that a divides b.
  - If $a \mid b$, then $\frac{b}{a}$ is an integer.
  - If a does not divide b, we write $a \nmid b$.

Properties of Divisibility

- Let a, b, and c be integers, where $a \neq 0$.
  - (i) If $a \mid b$ and $a \mid c$, then $a \mid (b + c)$.
  - (ii) If $a \mid b$, then $a \mid bc$ for all integers c.
  - (iii) If $a \mid b$ and $b \mid c$, then $a \mid c$.

- If a, b, and c are integers, where $a \neq 0$, such that $a \mid b$ and $a \mid c$, then $a \mid mb + nc$ whenever m and n are integers.

Division Algorithm

- Division
  - When an integer is divided by a positive integer, there is a quotient and a remainder.
  - This is called the division algorithm, but is actually a theorem.

- Division algorithm
  - If a is an integer and d a positive integer, then there are unique integers q and r, with $0 \leq r < d$, such that $a = dq + r$.
- d is called the *divisor*.
- a is called the *dividend*.
- q is called the *quotient*.
- r is called the *remainder*.

Definitions of functions

div and mod

$$q = a \text{ div } d$$

$$r = a \text{ mod } d$$

Division Algorithm

- Examples
  - When 101 is divided by 11,
    - The quotient is $9 = 101 \text{ div } 11$, and
    - The remainder is $2 = 101 \bmod 11$.
  - When -11 is divided by 3,
    - The quotient is $-4 = -11 \text{ div } 3$, and
    - The remainder is $1 = -11 \bmod 3$.

Congruence Relation

- Definition
  - If a and b are integers and m is a positive integer, then a is congruent to b modulo m if m divides $a - b$.
  - $a \equiv b \pmod{m}$
- We say that
  - $a \equiv b \pmod{m}$ is a congruence, and
  - m is the modulus of the congruence.
- Two integers are congruent $\bmod m$ if and only if they have the same remainder when divided by m.
- If a is not congruent to b modulo m, we write

$$a \not\equiv b \pmod{m}$$

Congruence Relation

- Example 1
  - Determine whether 17 is congruent to 5 modulo 6.
  - $17 \equiv 5 \pmod 6$ because 6 divides $17 - 5 = 12$.

- Example 2
  - Determine whether 24 and 14 are congruent modulo 6.
  - $24 \not\equiv 14 \pmod 6$ since $24 - 14 = 10$ is not divisible by 6.

$(\mathrm{mod}\ m)$ vs. $\mathrm{mod}\ m$

- The different use of mod:
  - $a \equiv b\ (\mathrm{mod}\ m)$
    - A relation on the set of integers
  - $a\ \mathrm{mod}\ m = b$
    - A function

- Theorem
  - Let a and b be integers, and let m be a positive integer.
  - Then $a \equiv b\ (\mathrm{mod}\ m)$ if and only if $a\ \mathrm{mod}\ m = b\ \mathrm{mod}\ m$.

Congruences of Sums and Products

- Theorem
  - Let m be a positive integer.
  - If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$.
- Proof
  - Because $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, there are integers s and t with $b = a + sm$ and $d = c + tm$.
  - Therefore,
    - $b + d = (a + sm) + (c + tm) = (a + c) + m(s + t)$
    - $bd = (a + sm)(c + tm) = ac + m(at + cs + stm)$
  - Hence, $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$.

Algebraic Manipulation of Congruences

- Multiplying both sides of a valid congruence by an integer preserves validity.
  - If $a \equiv b \pmod{m}$ holds, then $c \cdot a \equiv c \cdot b \pmod{m}$, where c is any integer.
- Adding an integer to both sides of a valid congruence preserves validity.
  - If $a \equiv b \pmod{m}$ holds then $c + a \equiv c + b \pmod{m}$, where c is any integer.
- Dividing a congruence by an integer does not always produce a valid congruence.
  - $14 \equiv 8 \pmod{6}$, but $7 \not\equiv 4 \pmod{6}$.
  - See page 33.

$\mathrm{mod}\ m$ Function of Products and Sums

- Let m be a positive integer and let a and b be integers.
  - $(a + b) \bmod m = \big((a \bmod m) + (b \bmod m)\big) \bmod m$
  - $ab \bmod m = \big((a \bmod m) \cdot (b \bmod m)\big) \bmod m$

Integer Representations

Representations of Integers

- Decimal (or base 10) notation
  - $965 = 9 \cdot 10^2 + 6 \cdot 10^1 + 5 \cdot 10^0$
- We can represent numbers using any base b, where b is a positive integer greater than 1.
- Important bases for computing and communications
  - $b = 2$ (*binary*)
  - $b = 8$ (*octal*)
  - $b = 16$ (*hexadecimal*)
- The ancient Mayans used base 20 and the ancient Babylonians used base 60.

Base b Representations

- Theorem
  - Let b be a positive integer greater than 1.
  - Then if n is a positive integer, it can be expressed

$$n = a_k b^k + a_{k-1} b^{k-1} + \cdots + a_1 b + a_0$$

where k is a nonnegative integer, a_0, a_1, …, a_k are nonnegative integers less than b, and $a_k \neq 0$.

  - The a_i, $i = 0, \ldots, k$ are called the base-b digits of the representation.

- Base b expansion
  - The above representation of n is called the base-b expansion of n and is denoted by $(a_k a_{k-1} \cdots a_1 a_0)_b$.
  - We usually omit the subscript 10 for base 10 expansions.

Base b Representations

- Binary expansions
  - Most computers represent integers and do arithmetic with binary (base 2) expansions of integers.
  - In these expansions, the only digits used are 0 and 1.
  - Example
    - What is the decimal expansion of the integer that has $(101011111)_2$ as its binary expansion?

- Octal expansions
  - The octal expansion (base 8) uses the digits $\{0,1,2,3,4,5,6,7\}$.
  - Example
    - What is the decimal expansion of the number with octal expansion $(111)_8$?

Base b Representations

- Hexadecimal Expansions
  - The hexadecimal expansion needs 16 digits, but our decimal system provides only 10.
  - So letters are used for the additional symbols.
  - The hexadecimal system uses the digits $\{0,1, \ldots, 9, A, B, \ldots, F\}$.
    - The letters A through F represent the decimal numbers 10~15.
- Example
  - What is the decimal expansion of the number with hexadecimal expansion $(2AE0B)_{16}$?
  - What is the decimal expansion of the number with hexadecimal expansion $(E5)_{16}$?

Base Conversion

- Base conversion algorithm

Procedure $base\_b(n, b$: positive integers with $b > 1)$

$q := n$

$k := 0$

while $(q \neq 0)$

$a_k := q \bmod b$

$q := q \operatorname{div} b$

$k := k + 1$

return $(a_{k-1}, a_{k-2}, \ldots, a_1, a_0)$

- q represents the quotient obtained by successive divisions by b, starting with $q = n$.
- The digits in the base b expansion are the remainders of the division given by $q \bmod b$.

Base Conversion

- Examples
  - Find the octal expansion of $(12345)_{10}$.
    - Successively dividing by 8 gives:

$$12345 = 8 \cdot 1543 + 1$$
$$1543 = 8 \cdot 192 + 7$$
$$192 = 8 \cdot 24 + 0$$
$$24 = 8 \cdot 3 + 0$$
$$3 = 8 \cdot 0 + 3$$

    - The remainders are the digits yielding $(30071)_8$.

  - Find the hexadecimal expansions of $(1111010111100)_2$.
    - Group the digits into blocks of four $(0011\ 1110\ 1011\ 1100)_2$ adding initial 0s as needed.
    - The blocks from left to right correspond to 3,E,B, and C.
    - Hence, the solution is $(3\mathrm{EBC})_{16}$.

Binary Addition of Integers

- Algorithm

Procedure $add(a, b$: positive integers$)$

binary expansions of $a := (a_{n-1}, a_{n-2}, \ldots, a_0)_2$

binary expansions of $b := (b_{n-1}, b_{n-2}, \ldots, b_0)_2$

$c := 0$

for $j := 1$ to $n - 1$

$d := \lfloor (a_j + b_j + c)/2 \rfloor$

$s_j := a_j + b_j + c - 2d$

$c := d$

$s_n := c$

return $(s_n, s_{n-1}, \ldots, s_1, s_0)_b$

- The number of additions of bits is $O(n)$.

Binary Multiplication of Integers

- Algorithm

Procedure $multiply(a, b$: positive integers$)$

  binary expansions of $a := (a_{n-1}, a_{n-2}, \ldots, a_0)_2$

  binary expansions of $b := (b_{n-1}, b_{n-2}, \ldots, b_0)_2$

  for $j := 0$ to $n - 1$

    if $b_j = 1$ **then** $c_j = a$ shifted j places.

    else $c_j := 0$

  $p := 0$

  for $j := 0$ to $n - 1$

    $p := p + c_j$

return p

- The number of additions of bits is $O(n^2)$.

Primes and Greatest Common Divisors

Primes

- Definition
  - A positive integer p greater than 1 is called prime if the only positive factors of p are 1 and p.
  - A positive integer that is greater than 1 and is not prime is called composite.

- Example
  - The integer 7 is prime because its only positive factors are 1 and 7, but 9 is composite because it is divisible by 3.

The Fundamental Theorem of Arithmetic

- Theorem
  - Every positive integer greater than 1 can be written uniquely as a prime or as the product of two or more primes where the prime factors are written in order of non-decreasing size.

- Examples
  - $100 = 2 \cdot 2 \cdot 5 \cdot 5 = 2^2 \cdot 5^2$
  - $641 = 641$
  - $999 = 3 \cdot 3 \cdot 3 \cdot 37 = 3^3 \cdot 37$
  - $1024 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^{10}$

Prime Factorization

Greatest Common Divisor

- Definition
  - Let a and b be integers, not both zero.
  - The largest integer d such that $d \mid a$ and also $d \mid b$ is called the greatest common divisor of a and b.
  - The greatest common divisor of a and b is denoted by $\gcd(a, b)$.
- Examples
  - What is the greatest common divisor of 24 and 36?
    - $\gcd(24, 36) = 12$
  - What is the greatest common divisor of 17 and 22?
    - $\gcd(17, 22) = 1$

Greatest Common Divisor

- The integers a and b are relatively prime if their GCD is 1.
  - Example: 17 and 22

- The integers $a_1, a_2, \ldots, a_n$ are pairwise relatively prime if $\gcd(a_i, a_j) = 1$ whenever $1 \leq i < j \leq n$.
  - Example: Determine whether the integers 10, 17 and 21 are pairwise relatively prime.
    - $\gcd(10, 17) = 1$, $\gcd(10, 21) = 1$, and $\gcd(17, 21) = 1$.
    - Hence 10, 17, and 21 are pairwise relatively prime.
  - Example: Determine whether the integers 10, 19, and 24 are pairwise relatively prime.
    - $\gcd(10,24) = 2$
    - 10, 19, and 24 are not pairwise relatively prime.

Finding GCD Using Prime Factorizations

- Suppose the prime factorizations of a and b are:

$$a = p_1^{a_1} \cdot p_2^{a_2} \cdots p_n^{a_n}, \qquad b = p_1^{b_1} \cdot p_2^{b_2} \cdots p_n^{b_n},$$

where each exponent is a non-negative integer. Then,

$$\gcd(a, b) = p_1^{\min(a_1, b_1)} \cdot p_2^{\min(a_2, b_2)} \cdots p_n^{\min(a_n, b_n)}.$$

- This formula is valid since the integer on the right divides both a and b. No larger integer can divide both a and b.

- Example
  - $120 = 2^3 \cdot 3 \cdot 5$ and $500 = 2^2 \cdot 5^5$
  - $\gcd(120, 500) = 2^{\min(3,2)} \cdot 3^{\min(1,0)} \cdot 5^{\min(1,3)} = 20.$

Finding GCD Using Euclidean Algorithm

- Euclidean algorithm for finding GCDs
  - An efficient method for computing the GCD of two integers.
  - Based on the idea that $\gcd(a, b) = \gcd(b, r)$ when $a > b$ and r is the remainder when a is divided by b.

- An example: Find $\gcd(91, 287)$.
  - $287 = 91 \cdot 3 + 14$
  - $91 = 14 \cdot 6 + 7$
  - $14 = 7 \cdot 2 + 0$ (Stopping condition: Found $r = 0$)
  - $\gcd(91, 287) = \gcd(91, 14) = \gcd(14, 7) = 7$

Finding GCD Using Euclidean Algorithm

- Pseudo-code

Procedure $gcd(a, b: \text{positive integers})$
  $x := a$
  $y := b$
  while $y \neq 0$
    $r := x \bmod y$
    $x := y$
    $y := r$
return x

Finding GCD Using Euclidean Algorithm

- Correctness of the algorithm
  - **Lemma**:
    Let $a = bq + r$, where a, b, q, and r are integers.
    Then $\gcd(a, b) = \gcd(b, r)$.
  - **Proof**:
    - Suppose that d divides both a and b.
      Then d also divides $a - bq = r$.
      Hence, any common divisor of a and b must also be a common divisor of b and r.
    - Suppose that d divides both b and r.
      Then d also divides $bq + r = a$.
      Hence, any common divisor of b and r must also be a common divisor of a and b.
    - Therefore, $\gcd(a, b) = \gcd(b, r)$.

Least Common Multiple

- Definition
  - The least common multiple of the positive integers a and b is the smallest positive integer that is divisible by both a and b.
  - It is denoted by $\mathrm{lcm}(a, b)$.

- The least common multiple can also be computed from the prime factorizations.

$$\mathrm{lcm}(a, b) = p_1^{\max(a_1, b_1)} \cdot p_2^{\max(a_2, b_2)} \cdots p_n^{\max(a_n, b_n)}.$$

  - This number is divided by both a and b, and no smaller number is divided by a and b.

- The GCD and the LCM of two integers are related by

$$ab = \gcd(a, b) \cdot \mathrm{lcm}(a, b)$$

GCD as Linear Combination

- Bézout's theorem
  - If a and b are positive integers, then there exist integers s and t such that $\gcd(a, b) = sa + tb$.

- Definition: Bézout's identity
  - If a and b are positive integers, then integers s and t such that $\gcd(a, b) = sa + tb$ are called Bézout coefficients of a and b .
  - The equation $\gcd(a, b) = sa + tb$ is called Bézout's identity.

- Example
  - $\gcd(6,14) = 2 = (-2) \cdot 6 + 1 \cdot 14$

Dividing Congruence by an Integer

- Dividing both sides of a valid congruence by an integer does not always produce a valid congruence.
- However, dividing by an integer relatively prime to the modulus does produce a valid congruence.

- Theorem
  - Let m be a positive integer and let a, b, and c be integers.
  - If $ac \equiv bc \pmod{m}$ and $\gcd(c, m) = 1$, then $a \equiv b \pmod{m}$.
- Proof
  - $ac \equiv bc \pmod{m} \Leftrightarrow m \mid (ac - bc) \Leftrightarrow m \mid c(a - b)$
  - Since $\gcd(c, m) = 1$, $m \mid (a - b)$.
  - Hence, $a \equiv b \pmod{m}$.

Solving Congruence

Linear Congruence

- Linear congruence
  - A congruence of the form $ax \equiv b \pmod{m}$, where m is a positive integer, a and b are integers, and x is a variable, is called a linear congruence.
  - The solutions to a linear congruence $ax \equiv b \pmod{m}$ are all integers x that satisfy the congruence.
- Inverse
  - An integer $\bar{a}$ such that $\bar{a}a \equiv 1 \pmod{m}$ is said to be "an inverse of a modulo m".
  - Example
    - 5 is an inverse of 3 modulo 7 since $5 \cdot 3 = 15 \equiv 1 \pmod{7}$.

One method of solving linear congruence makes use of an inverse $\bar{a}$, if it exists.
Although we cannot divide both sides of the congruence by a,
we can multiply by $\bar{a}$ to solve for x.

Inverse of a modulo m

- Theorem
  - If a and m are relatively prime integers and $m > 1$, then an inverse of a modulo m exists.
- Proof
  - [Bézout's Theorem] Since $\gcd(a, m) = 1$, there are integers s and t such that $sa + tm = 1$.
  - Hence, $sa + tm \equiv 1 \pmod{m}$.
  - Since $tm \equiv 0 \pmod{m}$, it follows that $sa \equiv 1 \pmod{m}$.
  - Consequently, s is an inverse of a modulo m.

The theorem guarantees that
an inverse of a modulo m exists
whenever a and m are relatively prime.

Finding Inverses

- The Euclidean algorithm and Bézout coefficients give us a systematic approach to finding inverses.
  - Example
    - Find an inverse of 3 modulo 7.
  - Solution
    - Because $\gcd(3,7) = 1$, an inverse of 3 modulo 7 exists. [by the theorem in the previous page]
    - Using the Euclidian algorithm: $7 = 2 \cdot 3 + 1$.
    - From this equation, we get $-2 \cdot 3 + 1 \cdot 7 = 1$, and see that −2 and 1 are Bézout coefficients of 3 and 7.
    - Hence −2 is an inverse of 3 modulo 7.
    - Also, every integer congruent to −2 modulo 7 is an inverse of 3 modulo 7 (*i.e.*, 5, −9, 12, …).

Finding Inverses

- The Euclidean algorithm and Bézout coefficients give us a systematic approach to finding inverses.
  - Example: Find an inverse of 101 modulo 4620.
  - Solution

$$4620 = 45 \cdot 101 + 75$$
$$101 = 1 \cdot 75 + 26$$
$$75 = 2 \cdot 26 + 23$$
$$26 = 1 \cdot 23 + 3$$
$$23 = 7 \cdot 3 + 2$$
$$3 = 1 \cdot 2 + 1$$
$$2 = 2 \cdot 1 + 0$$

$$\gcd(101, 4620) = 1$$

(1) Euclidean algorithm

$$1 = 3 - 1 \cdot 2$$
$$1 = 3 - 1 \cdot (23 - 7 \cdot 3) = -1 \cdot 23 + 8 \cdot 3$$
$$1 = -1 \cdot 23 + 8 \cdot (26 - 1 \cdot 23) = 8 \cdot 26 - 9 \cdot 23$$
$$1 = 8 \cdot 26 - 9 \cdot (75 - 2 \cdot 26) = 26 \cdot 26 - 9 \cdot 75$$
$$1 = 26 \cdot (101 - 1 \cdot 75) - 9 \cdot 75$$
$$= 26 \cdot 101 - 35 \cdot 75$$
$$1 = 26 \cdot 101 - 35 \cdot (4620 - 45 \cdot 101)$$
$$= -35 \cdot 4620 + 1601 \cdot 101$$

(2) Working backwards

Bézout coefficients = -35 and 1601 ➡ 1601 is an inverse of 101 modulo 4620.

Using Inverses to Solve Congruence

- What are the solutions of $3x \equiv 4 \pmod 7$?
  - We found that −2 is an inverse of 3 modulo 7 (two slides back).
  - We multiply both sides of the congruence by −2 giving
    $$-2 \cdot 3x \equiv -2 \cdot 4 \pmod 7$$
  - Because $-6 \equiv 1 \pmod 7$ and $-8 \equiv 6 \pmod 7$,
    it follows that if x is a solution, then $x \equiv -8 \equiv 6 \pmod 7$.
  - The solutions are the integers x such that $x \equiv 6 \pmod 7$,
    namely, $6, 13, 20, \ldots$ and $-1, -8, -15, \ldots$

[CSED261] Discrete Mathematics for Computer Science

Lecture 17

Relations

Suha Kwak

suha.kwak@postech.ac.kr

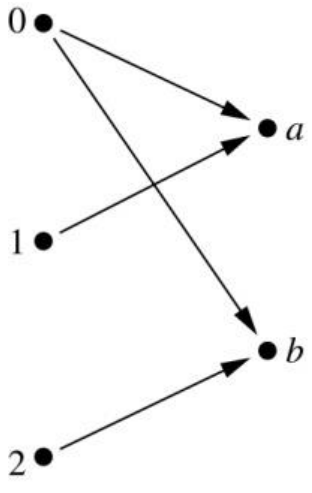
Dept. of Computer Science and Engineering

POSTECH

Relations and Their Properties

Binary Relation

- Definition
  - A binary relation R from a set A to a set B is a subset of $A \times B$.
- Example
  - Let $A = \{0,1,2\}$ and $B = \{a, b\}$.
  - $\{(0, a), (0, b), (1, a), (2, b)\}$ is a relation from A to B.
  - We can represent relations from a set A to a set B graphically or using a table:

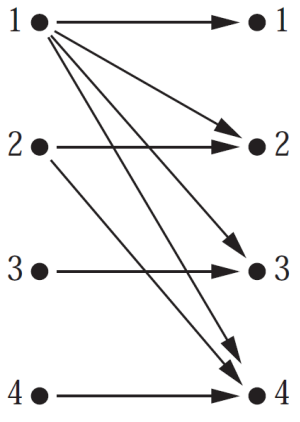


R	a	b
0	×	×
1	×	
2		×

Relations are more general than functions. A function is a relation where exactly one element of B is related to each element of A.

Binary Relation on a Set

- Definition
  - A binary relation R on a set A is a subset of $A \times A$ or a relation from A to A.
- Example 1
  - Suppose that $A = \{a, b, c\}$.
  - Then $R = \{(a, a), (a, b), (a, c)\}$ is a relation on A.
- Example 2
  - Let $A = \{1,2,3,4\}$.
  - The ordered pairs in the relation $R = \{(a, b) | a \text{ divides } b\}$ are $(1,1), (1,2), (1,3), (1,4), (2,2), (2,4), (3,3)$, and $(4,4)$.



R	1	2	3	4
1	×	×	×	×
2		×		×
3			×	
4				×

Binary Relation on a Set

- The number of binary relations on a set
  - Because a relation on A is a subset of $A \times A$, we count the subsets of $A \times A$.
  - Since $A \times A$ has n^2 elements when A has n elements, and a set with m elements has 2^m subsets, there are 2^{n^2} subsets of $A \times A$
  - Therefor, the number of binary relations on the set A is 2^{n^2}.

Binary Relation on a Set

- Example 3
  - Consider these relations on the set of integers:

$R_1 = \{(a,b)|a \leq b\}$, $R_4 = \{(a,b)|a = b\}$,
$R_2 = \{(a,b)|a > b\}$, $R_5 = \{(a,b)|a = b+1\}$,
$R_3 = \{(a,b)|a = b \text{ or } a = -b\}$, $R_6 = \{(a,b)|a+b \leq 3\}$.

These relations are on an infinite set and each of these relations is an infinite set.

  - Which of these relations contain each of the pairs $(1,1)$, $(1,2)$, $(2,1)$, $(1,-1)$, and $(2,2)$?

  - Solution:
    - $(1,1)$ is in R_1, R_3, R_4, and R_6.
    - $(1,2)$ is in R_1 and R_6.
    - $(2,1)$ is in R_2, R_5, and R_6.
    - $(1,-1)$ is in R_2, R_3, and R_6.
    - $(2,2)$ is in R_1, R_3, and R_4.

Reflexive Relations

- Definition
  - R is reflexive if and only if $(a, a) \in R$ for every element $a \in A$.
  - Written symbolically, R is reflexive if and only if
    $$\forall x[x \in A \rightarrow (x, x) \in R].$$
  - The empty relation on an empty set (*i.e.*, $A = \emptyset$) is reflexive.
- Example
  - The following relations on the integers are reflexive:
    - $R = \{(a, b) | a \leq b\}$
    - $R = \{(a, b) | a = b \text{ or } a = -b\}$
    - $R = \{(a, b) | a = b\}$
  - The following relations are not reflexive:
    - $R = \{(a, b) | a > b\}$
    - $R = \{(a, b) | a = b + 1\}$
    - $R = \{(a, b) | a + b \leq 3\}$

Symmetric Relations

- Definition
  - R is symmetric if and only if $(b, a) \in R$ whenever $(a, b) \in R$ for all $a, b \in A$.
  - Written symbolically, R is symmetric if and only if
$$\forall x \forall y \, [(x, y) \in R \rightarrow (y, x) \in R].$$
- Example
  - The following relations on the integers are symmetric:
    - $R = \{(a, b) | a = b \text{ or } a = -b\}$
    - $R = \{(a, b) | a = b\}$
    - $R = \{(a, b) | a + b \leq 3\}$
  - The following are not symmetric:
    - $R = \{(a, b) | a \leq b\}$
    - $R = \{(a, b) | a > b\}$
    - $R = \{(a, b) | a = b + 1\}$

Antisymmetric Relations

- Definition
  - A relation R on a set A such that $\forall a, b \in A$ if $(a, b) \in R$ and $(b, a) \in R$, then $a = b$ is called antisymmetric.
  - Written symbolically, R is antisymmetric if and only if
    $$\forall x \forall y \, [(x, y) \in R \wedge (y, x) \in R \rightarrow x = y].$$
- Example
  - The following relations on the integers are antisymmetric:
    - $R = \{(a, b) | a \leq b\}$
    - $R = \{(a, b) | a > b\}$
    - $R = \{(a, b) | a = b\}$
    - $R = \{(a, b) | a = b + 1\}$

For any integer,
if a $a \leq b$ and $b \leq a$,
then $a = b$.

  - The following relations are not antisymmetric:
    - $R = \{(a, b) | a = b \text{ or } a = -b\}$
    - $R = \{(a, b) | a + b \leq 3\}$

Transitive Relations

- Definition
  - A relation R on a set A is transitive if whenever $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$, for all $a, b, c \in A$.
  - Written symbolically, R is transitive if and only if $\forall x \forall y \forall z[(x, y) \in R \wedge (y, z) \in R \rightarrow (x, z) \in R\ ]$.
- Example
  - The following relations on the integers are transitive:
    - $R = \{(a, b) | a \leq b\}$
    - $R = \{(a, b) | a > b\}$
    - $R = \{(a, b) | a = b \text{ or } a = -b\}$
    - $R = \{(a, b) | a = b\}$
  - The following are not transitive:
    - $R = \{(a, b) | a = b + 1\}$
    - $R = \{(a, b) | a + b \leq 3\}$

For every integer,
If $a \leq b$ and $b \leq c$,
then $a \leq c$.

Combining Relations

- Given two relations R_1 and R_2, we can combine them using basic set operations to form new relations such as $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 - R_2$, and $R_2 - R_1$.

- Example
  - Let $A = \{1,2,3\}$ and $B = \{1,2,3,4\}$.
  - The relations $R_1 = \{(1,1), (2,2), (3,3)\}$ and $R_2 = \{(1,1), (1,2), (1,3), (1,4)\}$ can be combined using basic set operations to form new relations:

$$R_1 \cup R_2 = \{(1,1), (1,2), (1,3), (1,4), (2,2), (3,3)\}$$

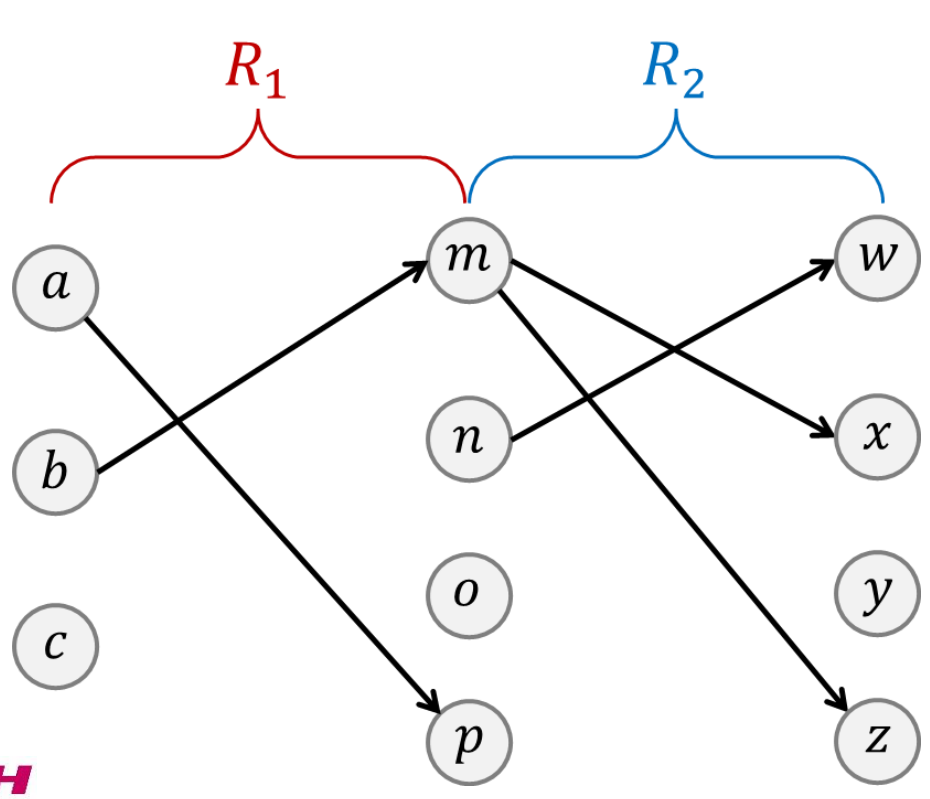
$$R_1 \cap R_2 = \{(1,1)\}$$

$$R_1 - R_2 = \{(2,2), (3,3)\}$$

$$R_2 - R_1 = \{(1,2), (1,3), (1,4)\}$$

Composition

- Definition
  - Suppose that R_1 is a relation from a set A to a set B and R_2 is a relation from B to a set C.
  - Then the composition (or composite) of R_2 with R_1 is a relation from A to C where if (x, y) is a member of R_1 and (y, z) is a member of R_2, then (x, z) is a member of $R_2 \circ R_1$.



$R_2 \circ R_1$
$= \{(b, x), (b, z)\}$

Powers of a Relation

- Definition
  - Let R be a binary relation on A. Then the powers R^n of the relation R can be defined inductively by:
    - Basis Step: $R^1 = R$
    - Inductive Step: $R^{n+1} = R^n \circ R$
- Theorem 1
  - The relation R on a set A is transitive if and only if $R^n \subseteq R$ for $n = 1,2,3, \ldots$.
  - In other words, the powers of a transitive relation are subsets of the relation.

Powers of a Relation

- Proof of Theorem 1: "If" part

The relation R on a set A is transitive if $R^n \subseteq R$ for $n = 1,2,3, \ldots$.

- Suppose that $R^n \subseteq R$ for $n = 1,2,3, \ldots$; in particular, $R^2 \subseteq R$.
- Note that if $(a, b) \in R$ and $(b, c) \in R$, then by the definition of composition, $(a, c) \in R^2$.
- Because $R^2 \subseteq R$, this means that $(a, c) \in R$.
- Hence, R is transitive.

Remind the definition of transitive relation

$\forall x \forall y \forall z[(x, y) \in R \wedge (y, z) \in R \rightarrow (x, z) \in R\ ]$

Powers of a Relation

- Proof of Theorem 1: "Only if" part (using math induction)

  $R^n \subseteq R$ for $n = 1,2,3, \ldots$ if the relation R on a set A is transitive.

  - Basis step: this part of the theorem is trivially true for $n = 1$.
  - Inductive step:
    - Assume that $R^n \subseteq R$, where n is a positive integer. We will show that this inductive hypothesis implies that $R^{n+1} \subseteq R$.
    - Assume that $(a, b) \in R^{n+1}$. Then, since $R^{n+1} = R^n \circ R$, there is an element $x \in A$ such that $(a, x) \in R$ and $(x, b) \in R^n$.
    - The inductive hypothesis $R^n \subseteq R$ implies that $(x, b) \in R$.
    - Furthermore, because R is transitive, and $(a, x) \in R$ and $(x, b) \in R$, it follows that $(a, b) \in R$.
    - We have shown that an arbitrary element (a, b) of R^{n+1} is included in R given the inductive hypothesis. Thus, $R^{n+1} \subseteq R$ when $R^n \subseteq R$.

Representing Relations

Representing Relations Using Matrices

- A relation between finite sets can be represented using a zero-one matrix.
  - Suppose R is a relation from $A = \{a_1, a_2, \dots, a_m\}$ to $B = \{b_1, b_2, \dots, b_n\}$.
  - The elements of the two sets can be listed in any particular arbitrary order. When $A = B$, we use the same ordering.
  - The relation R is represented by the matrix $M_R = [m_{ij}]$, where

$$m_{ij} = \begin{cases} 1, & \text{if } (a_i, b_j) \in R, \\ 0, & \text{if } (a_i, b_j) \notin R. \end{cases}$$

  - The matrix representing R has a 1 as its (i, j) entry when a_i is related to b_j and a 0 if a_i is not related to b_j.

Representing Relations Using Matrices

- Example 1
  - Suppose that $A = \{1,2,3\}$ and $B = \{1,2\}$. Let R be the relation from A to B containing (a, b) if $a \in A$, $b \in B$, and $a > b$.
  - Then what is the matrix representing R?
    Assume the ordering of elements is the same as the increasing numerical order.
  - Solution
    - Because $R = \{(2,1), (3,1), (3,2)\}$, the matrix is

$$M_R = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Representing Relations Using Matrices

- Example 2
  - Let $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3, b_4, b_5\}$. Which ordered pairs are in the relation R represented by the matrix below:

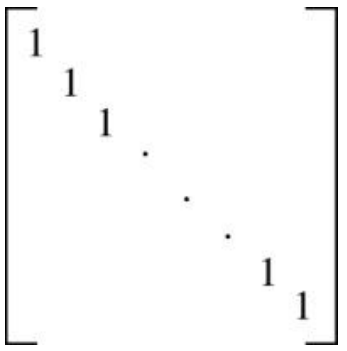
$$M_R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

- Solution
  - Because R consists of those ordered pairs (a_i, b_j) with $m_{ij} = 1$, it follows that:

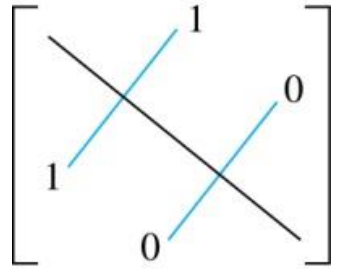
$$R = \{(a_1, b_2), (a_2, b_1), (a_2, b_3), (a_2, b_4), (a_3, b_1), (a_3, b_3), (a_3, b_5)\}.$$

Matrices of Relations on Sets

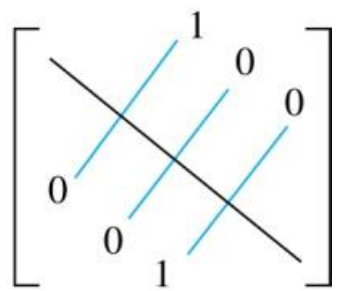
- If R is a reflexive relation, all the elements on the main diagonal of M_R are equal to 1.



- R is a symmetric relation if and only if $m_{ij} = 1$ whenever $m_{ji} = 1$.
- On the other hand, R is an antisymmetric relation, if and only if $m_{ij} = 0$ or $m_{ji} = 0$ when $i \neq j$.



(a) Symmetric



(b) Antisymmetric

Matrices of Relations on Sets

- Example 3
  - Suppose that the relation R on a set is represented by the matrix

$$M_R = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

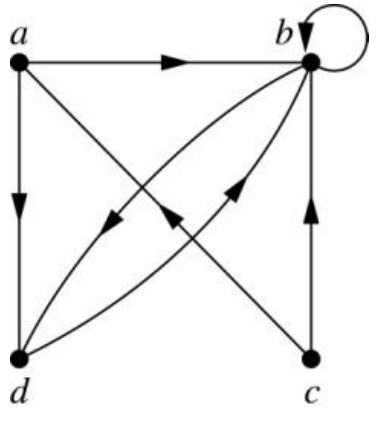
  - Is R reflexive, symmetric, and/or antisymmetric?

- Solution
  - Because all the diagonal elements are equal to 1, R is reflexive.
  - Because M_R is symmetric, R is symmetric.
  - Because both m_{12} and m_{21} are 1, M_R is not antisymmetric.

Representing Relations Using Digraphs

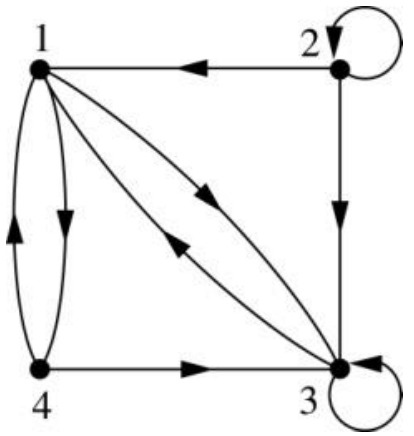
- Definition
  - A directed graph, or digraph, consists of a set V of vertices together with a set E of ordered pairs of elements of V called edges.
  - The vertex a is called the *initial vertex* of the edge (a, b), and the vertex b is called the *terminal vertex* of this edge.
  - An edge of the form (a, a) is called a loop.

- Example 1
  - The directed graph with vertices a, b, c, and d, and edges (a, b), (a, d), (b, b), (b, d), (c, a), (c, b), and (d, b) is



Representing Relations Using Digraphs

- Example 2
  - What are the ordered pairs in the relation represented by the directed graph below?

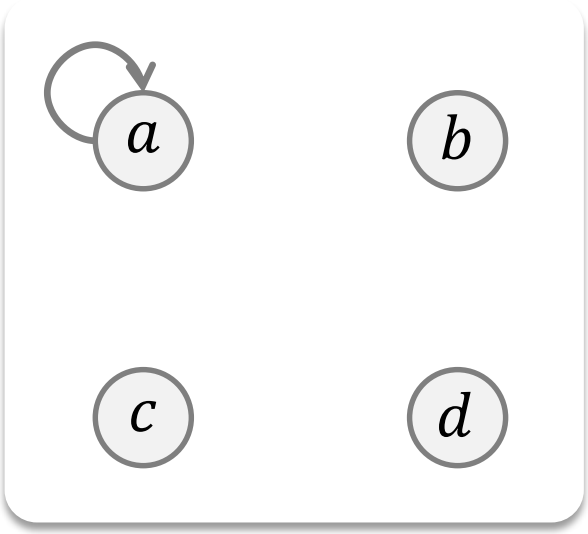


- Solution
  - The ordered pairs in the relation are $(1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 3), (4, 1)$, and $(4, 3)$.

Representing Relations Using Digraphs

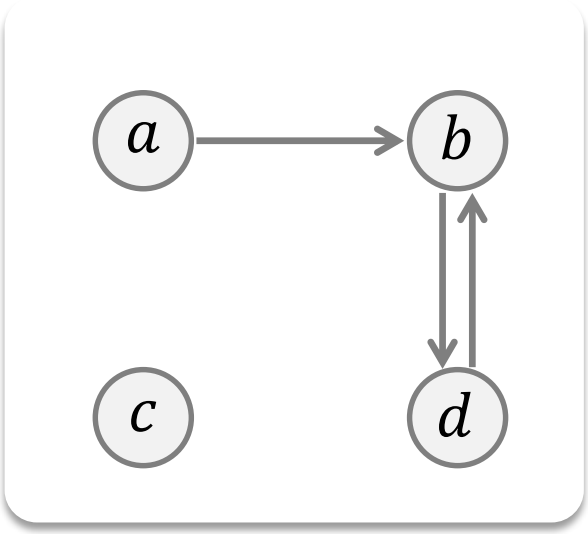
- Determining properties of a relation from its digraph
  - *Reflexivity*:
    A loop must be present at all vertices in the graph.
  - *Symmetry*:
    If (x, y) is an edge, then so is (y, x).
  - *Antisymmetry*:
    If (x, y) with $x \neq y$ is an edge, then (y, x) is not an edge.
  - *Transitivity*:
    If (x, y) and (y, z) are edges, then so is (x, z).

Representing Relations Using Digraphs



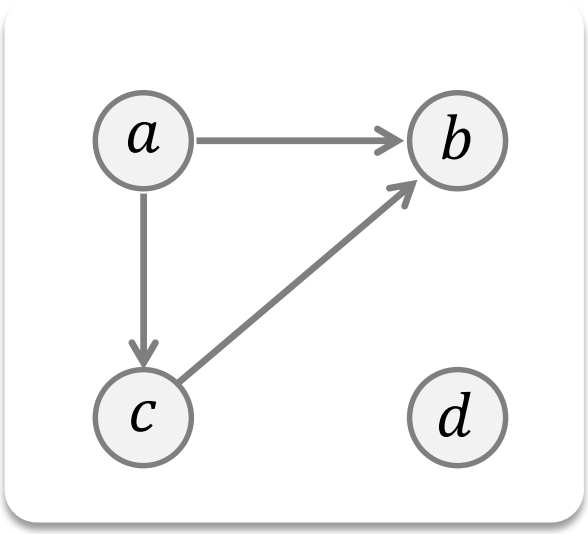
Reflexive?	No, not every vertex has a loop.
Symmetric?	Yes (trivially), there is no edge from one vertex to another.
Antisymmetric?	Yes (trivially), there is no edge from one vertex to another.
Transitive?	Yes (trivially), since there is no edge from one vertex to another.

Representing Relations Using Digraphs



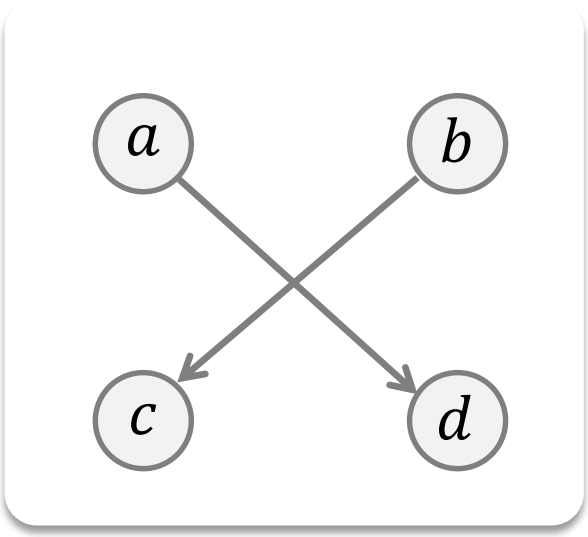
Reflexive?	No, there is no loop.
Symmetric?	No, there is an edge from a to b, but not from b to a.
Antisymmetric?	No, there is an edge from d to b and b to d.
Transitive?	No, there are edges from a to c and from c to b, but no edge from a to d.

Representing Relations Using Digraphs



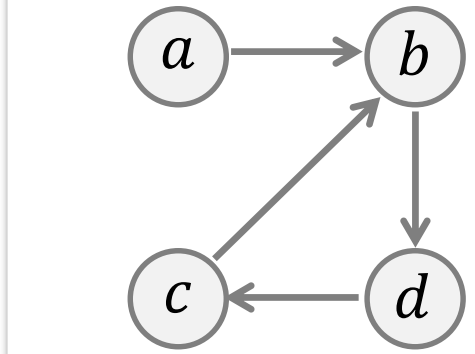
Reflexive?	No, there is no loop.
Symmetric?	No, for example, there is no edge from c to a.
Antisymmetric?	Yes, whenever there is an edge from one vertex to another, there is not one going back.
Transitive?	Yes.

Representing Relations Using Digraphs

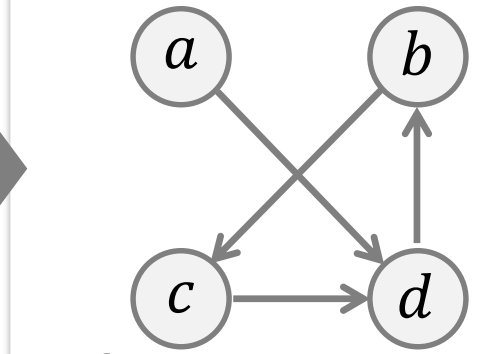


Reflexive?	No, there is no loop.
Symmetric?	No, for example, there is no edge from d to a.
Antisymmetric?	Yes, whenever there is an edge from one vertex to another, there is not one going back.
Transitive?	Yes (trivially), there are no two edges where the first edge ends at the vertex where the second begins.

Example of the Powers of a Relation

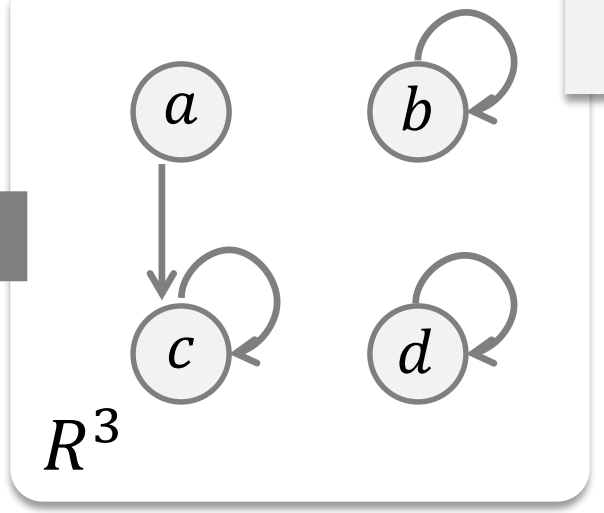


R

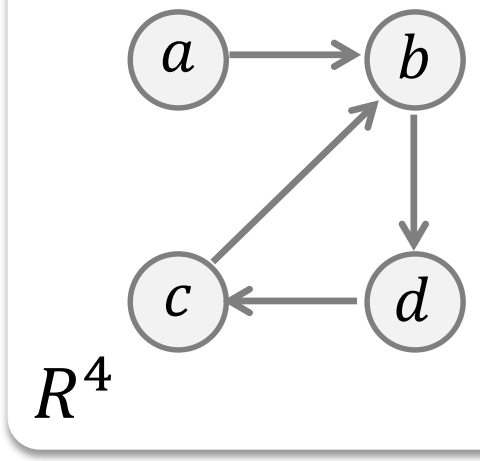


R^2

The pair (x, y) is in R^n if there is a path of length n from x to y in R (following the direction of the arrows).



R^3

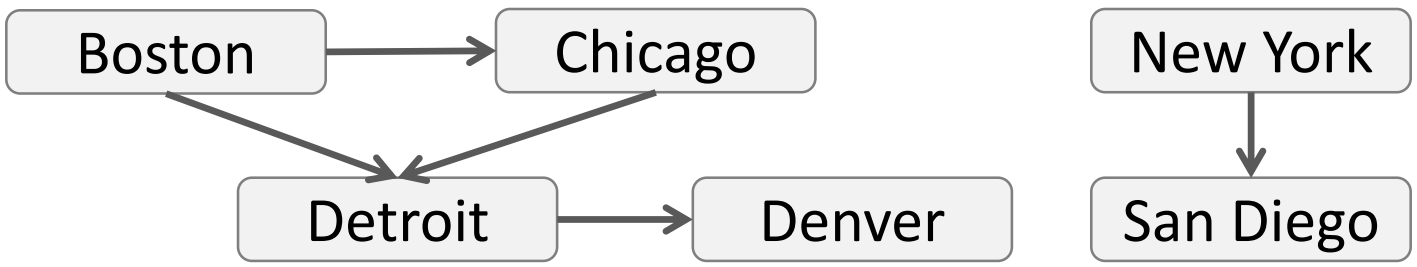


R^4

Closures of Relations

Closures

- Example
  - Direct, one-way telephone lines between data centers:



  - Let R be the relation containing (a, b) if there is a telephone line from the data center in a to that in b.
  - How can we determine if there is some (possibly indirect) link composed of one or more telephone lines from one center to another?
  - R cannot be used directly to answer this; as it is not transitive, it does not contain all the pairs that can be linked.

Closures

- Example (cont'd)
  - Instead, we can find all pairs of data centers that have a link by constructing a transitive relation S containing R such that S is a subset of every transitive relation containing R.
    (*i.e.*, S is the smallest transitive relation that contains R.)
  - This relation is called the transitive closure of R.

- A general description
  - Let R be a relation on a set A.
  - R may or may not have some property $\mathbf{P}$, such as reflexivity, symmetry, or transitivity.
  - If there is a relation S with property $\mathbf{P}$ containing R such that S is a subset of every relation with property $\mathbf{P}$ containing R, then S is called the closure of R with respect to $\mathbf{P}$.

Reflexive Closures

- How to produce the reflexive closure of R?
  - Given a relation R on a set A, the reflexive closure of R can be formed by adding to R all pairs of the form (a, a) with $a \in A$, not already in R.
  - The addition of these pairs produces a new relation that is reflexive, contains R, and is contained within any reflexive relation containing R.
  - We see that the reflexive closure of R equals $R \cup \Delta$, where $\Delta = \{(a, a) | a \in A\}$ is the *diagonal relation* on A.
- Example
  - What is the reflexive closure of the relation $R = \{(a, b) | a < b\}$ on the set of integers?
  - The reflexive closure of R is $R \cup \Delta = \{(a, b) | a < b\} \cup \{(a, a) | a \in \boldsymbol{Z}\} = \{(a, b) | a \leq b\}$.

Symmetric Closures

- How to produce the symmetric closure of R?
  - The symmetric closure of a relation R can be constructed by adding all ordered pairs of the form (b, a), where (a, b) is in the relation, that are not already present in R.
  - The symmetric closure can be constructed by taking the union of a relation with its inverse; that is, $R \cup R^{-1}$ is the symmetric closure of R, where $R^{-1} = \{(b, a) | (a, b) \in R\}$.

- Example
  - What is the symmetric closure of the relation $R = \{(a, b) | a > b\}$ on the set of positive integers?
  - The symmetric closure of R is $R \cup R^{-1} = \{(a, b) | a > b\} \cup \{(b, a) | a > b\} = \{(a, b) | a \neq b\}$.

Paths in Directed Graphs

- Review of a path in a directed graph
  - A *path* from a to b in a directed graph G is a sequence of edges $(x_0, x_1), (x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)$ in G, where n is a non-negative integer, and $x_0 = a$ and $x_n = b$.
  - This path is denoted by $x_0, x_1, x_2, x_3, \ldots, x_{n-1}, x_n$ and has length n.
  - We view the empty set of edges as a path of length zero from a to a.
  - A path of length $n \geq 1$ that begins and ends at the same vertex is called a *circuit* or *cycle*.

Paths in Directed Graphs

- The term *path* also applies to relations.
  - Carrying over the definition from directed graphs to relations, there is a path from a to b in R if there is a sequence of elements $a, x_1, x_2, x_3, \ldots, x_{n-1}, b$ with $(a, x_1) \in R$, $(x_1, x_2) \in R$, $\ldots$, and $(x_{n-1}, b) \in R$.

- Theorem
  - Let R be a relation on a set A.
  - There is a path of length n, where n is a positive integer, from a to b if and only if $(a, b) \in R^n$.
    (Refer to the "Example of the Powers of a Relation".)

Transitive Closures

- Finding the transitive closure of a relation is equivalent to determining which pairs of vertices in the associated directed graph are connected by a path.

- Definition: Connectivity relation
  - Let R be a relation on a set A.
  - The connectivity relation R^* consists of the pairs (a, b) such that there is a path of length at least one from a to b in R.
  - Because R^n consists of the pairs (a, b) such that there is a path of length n from a to b, it follows that R^* is the union of all the sets R^n, $R^* = \bigcup_{n=1}^{\infty} R^n$.

Transitive Closures

- Example of the connectivity relation
  - Let R be the relation on the set of all people in the world that contains (a, b) if a has met b.
  - What is R^n, where n is a positive integer greater than one?
    - The relation R^2 contains (a, b) if there is a person c such that $(a, c) \in R$ and $(c, b) \in R$, that is, if there is a person c such that a has met c and c has met b.
    - Similarly, R^n consists of those pairs (a, b) such that there are people $x_1, x_2, x_3, \ldots, x_{n-1}$ such that a has met x_1, x_1 has met x_2, $\ldots$, and x_{n-1} has met b.
  - What is R^*?
    - The relation R^* contains (a, b) if there is a sequence of people, starting with a and ending with b, such that each person in the sequence has met the next person in the sequence.

Transitive Closures

- Theorem
  - The transitive closure of a relation R equals the connectivity relation R^*.
  - Proof
    - R^* contains R by definition. To show that R^* is the transitive closure of R, we must also show that R^* is transitive and that $R^* \subseteq S$ whenever S is a transitive relation that contains R.
    - If $(a, b) \in R^*$ and $(b, c) \in R^*$, then there are (possibly indirect) paths from a to b and from b to c in R.
    - We obtain a path from a to c by starting with the path from a to b and following it with the path from b to c.
    - Hence, by $(a, c) \in R^*$. It follows that R^* is transitive.

Transitive Closures

- Theorem
  - The transitive closure of a relation R equals the connectivity relation R^*.
  - Proof (cont'd)
    - Suppose that S is a transitive relation containing R.
    - Because S is transitive, $S^n \subseteq S$.
      (See the proof of Theorem in p.13 of lecture note 17.)
    - Since $S^* = \bigcup_{k=1}^{\infty} S^k$ and $S^k \subseteq S$, it follows that $S^* \subseteq S$.
    - Now note that if $R \subseteq S$, then $R^* \subseteq S^*$, because any path in R is also a path in S.
    - Consequently, $R^* \subseteq S^* \subseteq S$.
    - Hence any transitive relation that contains R must also contain R^*. Therefore, R^* is the transitive closure of R.

Transitive Closures

- Lemma
  - Let A be a set with n elements, and let R be a relation on A.
  - If there is a path of length at least one in R from a to b, then there is such a path with length not exceeding n.
  - Moreover, when $a \neq b$, if there is a path of length at least one in R from a to b, then there is such a path with length not exceeding $n - 1$.

- From this lemma, we see that $R^* = \bigcup_{k=1}^{n} R^k$.
  - There is a path in R^* between two vertices if and only if there is a path between them in R^i for some positive integer $i \leq n$.
  - The zero-one matrix for the transitive closure (R^*) is the join of the zero-one matrices of the first n powers of the zero-one matrix of R. (See the theorem in the next page.)

Transitive Closures

- Theorem
  - Let M_R be the zero-one matrix of the relation R on a set with n elements.
  - Then the zero-one matrix of the transitive closure R^* is

$$M_{R*} = M_R \vee M_R^{[2]} \vee M_R^{[3]} \vee \cdots \vee M_R^{[n]}.$$

ALGORITHM 1 A Procedure for Computing the Transitive Closure.

procedure *transitive closure* ($\mathbf{M}_R$: zero–one $n \times n$ matrix)
$\mathbf{A} := \mathbf{M}_R$
$\mathbf{B} := \mathbf{A}$
for $i := 2$ **to** n
    $\mathbf{A} := \mathbf{A} \odot \mathbf{M}_R$
    $\mathbf{B} := \mathbf{B} \vee \mathbf{A}$
return $\mathbf{B}$ {$\mathbf{B}$ is the zero–one matrix for R^*}

Equivalence Relations

Equivalence Relations

- Definition 1
  - A relation on a set A is called an equivalence relation if it is reflexive, symmetric, and transitive.

- Definition 2
  - Two elements a and b that are related by an equivalence relation are called equivalent.
  - The notation $a \sim b$ is often used to denote that a and b are equivalent elements with respect to a particular equivalence relation.

Equivalence Relations

- Example
  - Suppose that R is the relation on the set of strings of English letters such that aRb if and only if $l(a) = l(b)$, where $l(x)$ is the length of the string x.
  - Is R an equivalence relation?

Two elements a and b are related by the relation R.

  - Solution
    - Show that all properties of an equivalence relation hold.
    - *Reflexivity*
      Because $l(a) = l(a)$, it follows that aRa for all strings a.
    - *Symmetry*
      Suppose that aRb. Since $l(a) = l(b)$, $l(b) = l(a)$ also holds and bRa.
    - *Transitivity*
      Suppose that aRb and bRc. Since $l(a) = l(b)$ and $l(b) = l(c)$, $l(a) = l(c)$ also holds and aRc.

Congruence Modulo m

- Example
  - Let m be an integer with $m > 1$.
  - Show that the relation $R = \{(a, b) | a \equiv b \pmod{m}\}$ is an equivalence relation on the set of integers.
  - Solution
    - Recall that $a \equiv b \pmod{m}$ if and only if m divides $a - b$.
    - *Reflexivity*
      $a \equiv a \pmod{m}$ since $a - a = 0$ is divisible by m since $0 = 0 \cdot m$.
    - *Symmetry*
      Suppose that $a \equiv b \pmod{m}$. Then $a - b$ is divisible by m, and so $a - b = km$, where k is an integer. It follows that $b - a = -km$, thus $b \equiv a \pmod{m}$.

Congruence Modulo m

- Example
  - Let m be an integer with $m > 1$.
  - Show that the relation $R = \{(a, b) | a \equiv b \pmod{m}\}$ is an equivalence relation on the set of integers.
  - Solution (cont'd)
    - *Transitivity*
      Suppose that $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$.
      Then m divides both $a - b$ and $b - c$, and there are integers k and l with $a - b = km$ and $b - c = lm$.
      Thus, $a - c = (a - b) + (b - c) = km + lm = (k + l)m$, which suggests that $a - c$ also divisible by m.
    - Since all the three properties hold, congruence module m is an equivalence relation.

Divides

- Example
  - Show that the “divides” relation on the set of positive integers is not an equivalence relation.
  - Solution
    - The reflexivity, and transitivity do hold, but symmetry does not. Hence, “divides” is not an equivalence relation.
    - *Reflexivity*
      $a \mid a$ for all a.
    - *Not Symmetric*
      A counter-example: $2 \mid 4$, but $4 \nmid 2$.
    - *Transitivity*
      Suppose that a divides b and b divides c. Then there are positive integers k and l such that $b = ak$ and $c = bl$. Hence, $c = akl$, so a divides c. Therefore, the relation is transitive.

Equivalence Classes

- Definition 3
  - Let R be an equivalence relation on a set A.
  - The set of all elements that are related to an element a of A is called the equivalence class of a.
  - The equivalence class of a with respect to R is denoted by $[a]_R$.
    - $[a]_R = \{s|(a,s) \in R\}$.
    - When only one relation is under consideration, we can write $[a]$, without the subscript R, for this equivalence class.
  - If $b \in [a]_R$, then b is called a representative of this equivalence class; any element of a class can be used as a representative of the class.

Equivalence Classes

- Congruence class modulo m
  - The equivalence classes of the relation "*congruence modulo* m" are called the "*congruence classes modulo* m".
  - The congruence class of an integer a modulo m is denoted by $[a]_m$; $[a]_m = \{\dots, a-2m, a-m, a, a+m, a+2m, \dots\}$.

- Example
  - $[0]_4 = \{\dots, -8, -4, 0, 4, 8, \dots\}$
  - $[1]_4 = \{\dots, -7, -3, 1, 5, 9, \dots\}$
  - $[2]_4 = \{\dots, -6, -2, 2, 6, 10, \dots\}$
  - $[3]_4 = \{\dots, -5, -1, 3, 7, 11, \dots\}$

Equivalence Classes and Partitions

- Theorem 1
  - Let R be an equivalence relation on a set A. Then these statements for elements a and b of A are equivalent:

    (*i*) aRb

    (*ii*) $[a] = [b]$

    (*iii*) $[a] \cap [b] \neq \emptyset$

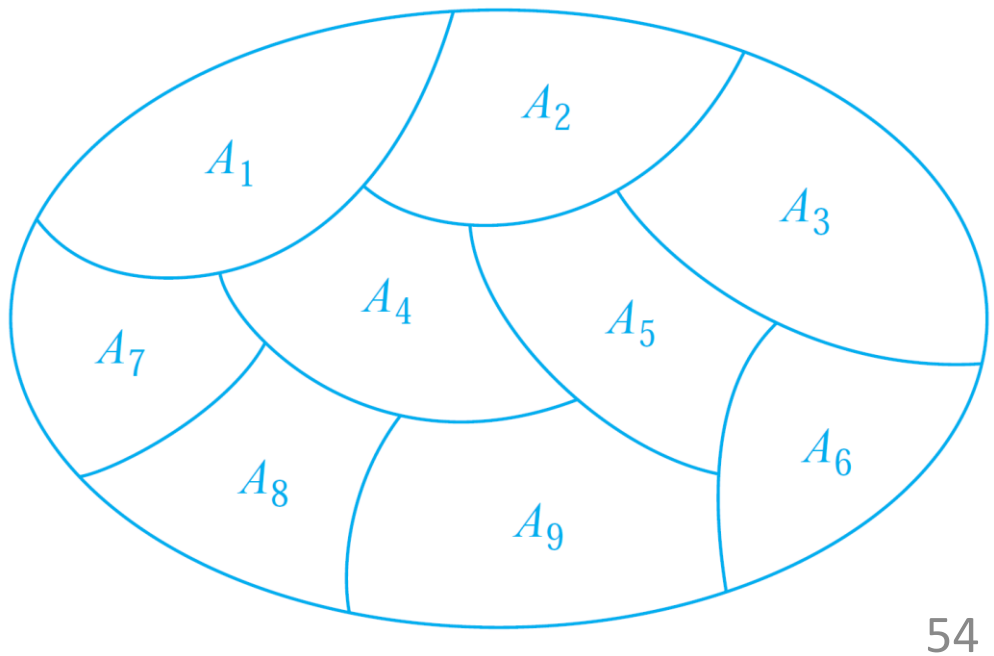
  - Proof: (*i*) implies (*ii*).
    - Assume that aRb and that $c \in [a]$. Then aRc.
    - Because aRb and R is symmetric, bRa.
    - Because R is transitive and bRa and aRc, it follows that bRc.
    - Hence, $c \in [b]$. Therefore, $[a] \subseteq [b]$.
    - A similar argument shows that $[b] \subseteq [a]$.
    - Since $[a] \subseteq [b]$ and $[b] \subseteq [a]$, we have shown that [a] = [b].

Equivalence Classes and Partitions

- Theorem 1
  - Let R be an equivalence relation on a set A. Then these statements for elements a and b of A are equivalent:
    (i) aRb
    (ii) $[a] = [b]$
    (iii) $[a] \cap [b] \neq \emptyset$
  - Proof: (*ii*) implies (*iii*).
    - Assume that $[a] = [b]$. It follows that $[a] \cap [b] \neq \emptyset$ because $[a]$ is nonempty (since R is reflexive, $a \in [a]$).
  - Proof: (*iii*) implies (*i*).
    - Suppose that $[a] \cap [b] \neq \emptyset$. Then there is an element c with $c \in [a]$ and $c \in [b]$. In other words, aRc and bRc.
    - By the symmetric property of R, cRb. Then by transitivity, because aRc and cRb, we have aRb.

Partition of a Set

- Definition
  - A partition of a set S is a collection of disjoint nonempty subsets of S that have S as their union.
  - In other words, the collection of subsets A_i, where $i \in I$ (I: a set of indices), forms a partition of S if and only if
    - $A_i \neq \emptyset$ for $i \in I$,
    - $A_i \cap A_j = \emptyset$ when $i \neq j$, and
    - $\bigcup_{i \in I} A_i = S$.



An Equivalence Relation Partitions a Set

- Equivalence relation partitioning a set
  - Let R be an equivalence relation on a set A.
  - The union of all the equivalence classes of R is all of A, since an element a of A is in its own equivalence class $[a]_R$.
    - In other words, $\bigcup_{a \in A} [a]_R = A$.
  - From Theorem 1, it follows that these equivalence classes are either equal or disjoint, so $[a]_R \cap [b]_R = \emptyset$ when $[a]_R \neq [b]_R$.
  - Therefore, the equivalence classes form a partition of A, because they split A into disjoint subsets.

Theorem 1. The following statements are equivalent:
(*i*) aRb
(*ii*) $[a] = [b]$
(*iii*) $[a] \cap [b] \neq \emptyset$

An Equivalence Relation Partitions a Set

- Theorem 2
  - Let R be an equivalence relation on a set S. Then the equivalence classes of R form a partition of S.
  - Conversely, given a partition $\{A_i | i \in I\}$ of the set S, there is an equivalence relation R that has the sets A_i, $i \in I$, as its equivalence classes.
  - Proof
    - We have already shown the first part of the theorem.
    - For the second part, assume that $\{A_i | i \in I\}$ is a partition of S.
    - Let R be the relation on S consisting of the pairs (x, y) where x and y belong to the same subset A_i in the partition.
    - We must show that R satisfies the properties of an equivalence relation.

An Equivalence Relation Partitions a Set

- Theorem 2
  - Let R be an equivalence relation on a set S. Then the equivalence classes of R form a partition of S.
  - Conversely, given a partition $\{A_i | i \in I\}$ of the set S, there is an equivalence relation R that has the sets A_i, $i \in I$, as its equivalence classes.
  - Proof (cont'd)
    - *Reflexivity*: For every $a \in S$, $(a, a) \in R$ because a is in the same subset as itself.
    - *Symmetry*: If $(a, b) \in R$, then b and a are in the same subset of the partition, so $(b, a) \in R$.
    - *Transitivity*: If $(a, b) \in R$ and $(b, c) \in R$, then a and b are in the same subset of the partition, as are b and c. Since the subsets are disjoint and b belongs to both, the two subsets of the partition must be identical. Therefore, $(a, c) \in R$ since a and c belong to the same subset of the partition.

[CSED261] Discrete Mathematics for Computer Science

Lecture 16

Discrete Probability

Suha Kwak

suha.kwak@postech.ac.kr

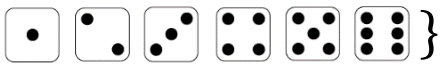
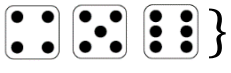
Dept. of Computer Science and Engineering

POSTECH

An Introduction to Discrete Probability

Probability of an Event

- 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.

- Example
  - Random experiment
    - Tossing a single six-sided dice
  - Sample space
    - $S = \{$  $\}$
  - Event
    - A subset of the sample space
    - "The number on the face > 3" $= \{$  $\}$
    - $S =$ Sure event
    - $\phi =$ Impossible event

Probability of an Event

- Definition (*by Pierre-Simon Laplace*)
  - 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

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

  - 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 $0 \leq |E| \leq |S|$.

- 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?
  - Answer: $\frac{4}{9}$

Probability of an Event

- Example 2
  - 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?
  - Answer:
    - 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 $\frac{1}{3{,}838{,}380} \approx 0.00000026.$

Probability of an Event

- Example 3
  - 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.
  - Answer:
    - Use the product rule in each case.
    A. (*Sampling without replacement*) The probability is $\frac{1}{254,251,200}$ 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*) Since $50^5 = 312,500,000$ the probability is $\frac{1}{50^5} = \frac{1}{312,500,000}$.

Complements and Unions of Events

- Theorem: Probability of the complementary event
  - 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 $|\bar{E}| = |S| - |E|$,

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

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

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

Complements and Unions of Events

- Theorem: Probability of a union of events
  - 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,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

    - It follows that

$$P(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|}$$

$$= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Complements and Unions of Events

- Example
  - Suppose we select an integer at random from the set of positive integers not exceeding 100.
  - What is the probability that the selected integer 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

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

$$= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}$$

Probability Theory

Summary

- Assigning Probabilities
- Conditional Probability
- Independence
- Bernoulli Trials and the Binomial Distribution
- Random Variables

Assigning Probabilities

- Laplace's definition of probability
  - It assumes that all outcomes are equally likely.
  - A more general definition is required to avoid the restriction.

- Assigning probabilities
  - Let S be a sample space of an experiment with a finite number of outcomes.
  - Assign a probability $P(s)$ to each outcome s, so that:
    - $0 \leq P(s) \leq 1$ for each $s \in S$
    - $\sum_{s \in S} P(s) = 1$
  - 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:
    - For a fair coin, the probabilities for *H* and *T* are both $\frac{1}{2}$.
    - For the biased coin: $P(H) = 2 \cdot P(T)$, and since $P(H) + P(T) = 1$, $P(H) = \frac{2}{3}$ and $P(T) = \frac{1}{3}$.

Uniform Distribution

- Definition
  - Suppose that S is a set with n elements.
  - The uniform distribution assigns the probability $\frac{1}{n}$ to each element of S.
  - Note that we could have used Laplace's definition here.

- Example
  - For a fair coin: $P(H) = P(T) = \frac{1}{2}$
  - For a fair six-faced dice: $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$

Probability of an Event

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

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

  - Now no assumption is being made about the distribution.

- Example
  - Suppose that a die is biased so that ⚂ appears twice as often as each other, but the other five faces appear equally.
  - What is the probability that an odd number appears when we roll this die?

$P(⚂) = \frac{2}{7}$, and $P(⚀) = P(⚁) = P(⚃) = P(⚄) = P(⚅) = \frac{1}{7}$

$P(E) = P(⚀⚂⚄) = P(⚀) + P(⚂) + P(⚄) = \frac{4}{7}$

Combinations of Events

- Theorem
  - If $E_1, E_2, \ldots$ 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 4 is sampled from a uniform distribution.
  - What is the probability that it contains at least two consecutive 0s, *given* that its first bit is a 0?
    - 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.
    - $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, so $P(E \cap F) = 5/16$.
    - $P(F) = 1/2$ since there are 8 bit strings starting with a 0.
    - Hence, $P(E|F) = 5/16 \cdot 2 = 5/8$

Independence

- Definition
  - The events E and F are independent if and only if
$$P(E \cap F) = P(E) \cdot P(F)$$
- Example
  - Suppose E is the event that a randomly generated bit string of length 4 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 4 are equally likely?
  - They are independent; why?
    - There are eight bit strings of length 4 beginning with a 1, and eight bit strings of length 4 that contain an even number of 1s.
    - Since the number of bit strings of length 4 is 16, $P(E) = P(F) = \frac{1}{2}$
    - Since $E \cap F = \{1111,1100,1010,1001\}$, $P(E \cap F) = \frac{1}{4}$.

Pairwise and Mutual Independence

- Definition (*Pairwise independence)*
  - The events $E_1, E_2, \ldots, E_n$ are pairwise independent if and only if

$$P(E_i \cap E_j) = P(E_i) \cdot P(E_j)$$

for all pairs i and j with $1 \leq i < j \leq n$.

- Definition (*Mutual independence*)
  - The events $E_1, E_2, \ldots, E_n$ are mutually independent if and only if

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

whenever i_j $(j = 1,2, \ldots, m)$ are integers with

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

Pairwise and Mutual Independence

- Example
  - A fair 6-sided dice is thrown twice.
  - Three events of our interest:
    - E_1: The first roll is a 3.
    - E_2: The second roll is 4.
    - E_3: The sum of the two rolls is 7.
  - The three events are pairwise independent since
    - $P(E_1) = P(E_2) = P(E_3) = \frac{1}{6}$
    - $P(E_1 \cap E_2) = P(E_2 \cap E_3) = P(E_1 \cap E_3) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$
  - However, the events are not mutually independent since
    - $P(E_1 \cap E_2 \cap E_3) = \frac{1}{36} \neq \left(\frac{1}{6}\right)^3$

Bernoulli Trials

- 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 4 heads occur when the coin is flipped 7 times?
  - Solution:
    - There are $2^7 = 128$ possible outcomes.
    - The number of ways 4 of the 7 flips can be heads is $C(7,4)$.
    - The probability of each of the outcomes is $\left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^3$ since the 7 flips are independent.
    - Hence, the probability that exactly 4 heads occur is

$$C(7,4) \cdot \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^3 = \frac{35 \cdot 16}{3^7} = \frac{560}{2187}$$

Bernoulli Trials

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

$$C(n, k) \cdot p^k \cdot q^{n-k}$$

  - Proof
    - The outcome of n Bernoulli trials is an n-tuple $(t_1, t_2, \dots, t_n)$, where each t_i is 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 $p^k q^{n-k}$.
    - Because there are a $C(n, k)$ number of n-tuple of Ss and Fs that contain exactly k Ss, the probability of k successes is $C(n, k) \cdot p^k \cdot q^{n-k}$.
  - We denote this probability by $b(k: n, p) = C(n, k) \cdot p^k \cdot q^{n-k}$, a function of k given n and p.

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!

In the late 1940s W. Feller and J.L. Doob flipped a coin to see whether both would use “random variable” or the more fitting “chance variable.” Unfortunately, Feller won and the term “random variable” has been used ever since…

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: A coin flipped 3 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 8 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

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:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

  - Proof

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E \cap F)}{P(E \cap F) + P(E \cap \bar{F})}$$

$$= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

Bayes' Theorem

- Example
  - We have two boxes.
    - The first box contains 2 green balls and 7 red balls.
    - The second contains 4 green balls and 3 red balls.
  - You select one of the boxes at random, then select a ball from the selected box at random.
  - If you have a red ball, what is the probability that you selected a ball from the first box.
  - Solution
    - E: The event that you have chosen a red ball
    - F: The event that you have chosen the first box

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

$$= \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{49}{76}$$

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

$$\bigcup_{i=1}^{n} F_i = S$$

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

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

A Little Taste of Machine Learning

- Interpreting Bayes' theorem

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

  - The event of our interest A
    - Typically latent (*i.e.*, hidden)
  - The event as an observation B
  - Prior probability $P(A)$
    - Probability of the event A with no observation
    - Based only on our *prior* knowledge about A
  - Likelihood $P(B|A)$
    - Probability of observing B when A happens
  - Posterior probability $P(A|B)$
    - Probability of A if we observed B

$$posterior \propto likelihood \times prior$$

A Little Taste of Machine Learning

- Interpreting Bayes' theorem (cont'd)

x5 x3 x2

$[prior]$

$P(\text{🍎}) = 0.5, \quad P(\text{🍐}) = 0.3, \quad P(\text{🍊}) = 0.2$

$[likelihood]$

$P(\text{👅} \mid \text{🍎}) = 0.3, \quad P(\text{👅} \mid \text{🍐}) = 0.5, \quad P(\text{👅} \mid \text{🍊}) = 0.9$

$[posterior \propto likelihood \times prior]$

$P(\text{🍎} \mid \text{👅}) \propto 0.15, \quad P(\text{🍐} \mid \text{👅}) \propto 0.15, \quad P(\text{🍊} \mid \text{👅}) \propto 0.18$

A Little Taste of Machine Learning

- Bayesian spam filtering
  - Suppose that we have an initial set B of spam messages and set G of non-spam messages.
    - Datasets like B and G are called *training data* in the context of machine learning
  - We can use this information along with Bayes' theorem to predict the probability that a new email message is spam.
  - We look at a particular word w, and count the number of times that it occurs in B and in G: $n_B(w)$ and $n_G(w)$.
  - Let S be the event that the message is spam, and E be the event that the message contains the word w.
    - Estimated probability that a spam message contains w:
      $$P(E|S) = n_B(w)/|B|$$
    - Estimated probability that a non-spam message contains w:
      $$P(E|\bar{S}) = n_G(w)/|G|$$

A Little Taste of Machine Learning

- Bayesian spam filtering (cont'd)
  - For spam filtering, we compute the posterior probability $P(S|E)$ for the message and categorize it as spam if the probability is larger than a pre-defined threshold.
  - $P(S|E)$ indicates the probability that the message is spam when it contains the word w.

$$P(S|E) = \frac{P(E|S)P(S)}{P(E)} = \frac{P(E|S)P(S)}{P(E|S)P(S) + P(E|\bar{S})P(\bar{S})}$$

$$= \frac{P(E|S)}{P(E|S) + P(E|\bar{S})}$$

Here we assuming that the prior probabilities are uniform.

If we have data on the frequency of spam messages, we can obtain a better estimate of the priors.

A Little Taste of Machine Learning

- Bayesian spam filtering: A specific example
  - We find that the word “Rolex” occurs in 250 out of 2000 spam messages and occurs in 5 out of 1000 non-spam messages.
  - Estimate the probability that an incoming message with “Rolex” is spam.
  - Suppose our threshold for rejecting the email is 0.9.
  - Is the message a spam according to our Bayesian spam filter?

$$P(\text{"Rolex"}|S) = \frac{250}{2000} = 0.125, \qquad P(\text{"Rolex"}|\bar{S}) = \frac{5}{1000} = 0.005$$

$$P(S|\text{"Rolex"}) = \frac{P(\text{"Rolex"}|S)}{P(\text{"Rolex"}|S) + P(\text{"Rolex"}|\bar{S})} = \frac{0.125}{0.125 + 0.005} \approx 0.962$$

Since the probability $P(S|\text{"Rolex"})$ is larger than the threshold,
our spam filtering system will categorize the message as spam and reject it!

A Little Taste of Machine Learning

- Bayesian spam filtering with using multiple words
  - Accuracy can be improved by considering more than one word as evidence.
  - Consider the case where E_1 and E_2 denote the events that the message contains the words w_1 and w_2 respectively.
  - For simplicity, we make the assumption that $P(S) = 0.5$ (the uniform prior) and E_1 and E_2 are *conditionally independent* given S:

$$P(E_1 \cap E_2|S) = P(E_1|S)P(E_2|S)$$

- Then the posterior probability of S is given by

$$P(S|E_1 \cap E_2) = \frac{P(E_1|S)P(E_2|S)}{P(E_1|S)P(E_2|S) + P(E_1|\bar{S})P(E_2|\bar{S})}$$

A Little Taste of Machine Learning

- Bayesian spam filtering with using multiple words:
  A specific example
  - We have 2000 spam messages and 1000 non-spam messages.
  - The word “stock” occurs 400 times in the spam messages and 60 times in the non-spam.
  - The word “undervalued” occurs in 200 spam messages and 25 non-spam.

$$P(E_1|S) = \frac{400}{2000} = 0.2, \quad P(E_1|\bar{S}) = \frac{60}{1000} = 0.06$$

$$P(E_2|S) = \frac{200}{2000} = 0.1, \quad P(E_2|\bar{S}) = \frac{25}{1000} = 0.025$$

$$P(S|E_1 \cap E_2) = \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.93$$

A Little Taste of Machine Learning

- Bayesian spam filtering in general
  - The more words we consider, the more accurate the spam filter we can obtain.
  - With the independence assumption if we consider k words:

$$P\left(S\middle|\bigcap_{i=1}^{k} E_i\right) = \frac{\prod_{i=1}^{k} P(E_i|S)}{\prod_{i=1}^{k} P(E_i|S) + \prod_{i=1}^{k} P(E_i|\bar{S})}$$

  - We can further improve the filter by considering pairs of words as a single block of certain types of strings.

Expected Value and Variance

Expected Value

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

$$E(X) = \sum_{s \in S} P(s) \cdot X(s)$$

- Example
  - 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 2
  - Suppose that n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial.
  - Then the expected number of successes in this case is np.
  - Proof
    - Let X be the random variable equal to the number of success in n trials.
    - Then,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

    - Hence,

$$E(X) = \sum_{k=1}^{n} k \cdot P(X = k)$$

$$= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Expected Value

- Theorem 2
  - Proof (cont'd)

$$E(X) = \sum_{k=1}^{n} k \cdot P(X = k) = \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

Shift the index of sum with $j = k - 1$.

$$= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

Recall the binomial theorem:

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j$$

$$= np \cdot (p + 1 - p)^{n-1} = np \cdot 1 = np$$

Linearity of Expectations

- Theorem
  - X_i are random variables on S where $i = 1,2, \dots, n$
  - 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) $E(aX + b) = aE(X) + b$

- Proof of (i)
  - When $n = 2$,

$$E(X_1 + X_2) = \sum_{s \in S} P(s)\big(X_1(s) + X_2(s)\big)$$

$$= \sum_{s \in S} P(s)X_1(s) + \sum_{s \in S} P(s)X_2(s) = E(X_1) + E(X_2)$$

  - We can generalize the result for $n > 2$ by math induction.

Linearity of Expectations

- Theorem
  - X_i are random variables on S where $i = 1,2, \dots, n$
  - 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) $E(aX + b) = aE(X) + b$

- Proof of (ii)

$$E(aX + b) = \sum_{s \in S} P(s)(aX(s) + b)$$

$$= a \sum_{s \in S} P(s)X(s) + b \sum_{s \in S} P(s)$$

$$= aE(X) + b \cdot 1 = aE(X) + b$$

Average-Case Computational Complexity

- The average-case complexity of an algorithm can be found by computing the expected value of a random variable.
  - Let the sample space of an experiment be the set of possible inputs a_j where $j = 1,2, \dots, n$.
  - Let the random variable X be the assignment to a_j of the number of operations used by the algorithm when given a_j as input.
  - Assign a probability $P(a_j)$ to each possible input value a_j.
  - The expected value of X is the average-case computational complexity of the algorithm.

$$E(X) = \sum_{j=1}^{n} P(a_j) X(a_j)$$

Average-Case Computational Complexity

- Linear search

Procedure linear_search(x: integer, $\{a_1, a_2, \ldots, a_n\}$: integers)

$i := 1$

while $i \leq n$ and $x \neq a_i$

$i := i + 1$

if $i \leq n$ **then** $\ell := i$

else $\ell := 0$

return ℓ

- Suppose that
  - The probability that x is in the list is p,
  - It is equally likely that x is any of the n elements of the list.
- Then, what is the average-case computational complexity of this algorithm?

Average-Case Computational Complexity

- Linear search (cont'd)
  - There are $n + 1$ possible types of input:
    - One type for each of the n numbers on the list
    - One additional type for the numbers not on the list.
  - Recall that we need:
    - $2i + 1$ comparisons if x equals the i^{th} element of the list.
    - $2n + 2$ comparisons if x is not on the list.
  - The probability that x equals a_j is p/n and the probability that x is not in the list is $q = 1 - p$.
  - The average-case computational complexity of the linear search algorithm is:

$$\begin{aligned} E = {} & 3p/n + 5p/n + \cdots + (2n+1)p/n + (2n+2)q \\ = {} & (p/n)\big(3 + 5 + \cdots + (2n+1)\big) + (2n+2)q \\ = {} & (p/n)\big((n+1)^2 - 1\big) + (2n+2)q \\ = {} & p(n+2) + (2n+2)q. \end{aligned}$$

For Independent Random Variables

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

$$P(X = r_1 \text{ and } Y = r_2) = P(X = r_1) \cdot P(Y = r_2)$$

- Theorem
  - If X and Y are independent variables on a sample space S, then $E(XY) = E(X) \cdot E(Y)$.
- Proof
  - The event $XY = r$ is the disjoint union of the events $X = r_1$ and $Y = r_2$ over all $r_1 \in X(S)$ and $r_2 \in Y(S)$ with $r = r_1 \cdot r_2$.

$$E(XY) = \sum_{r \in XY(S)} r \cdot P(XY = r)$$

$$= \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1 \text{ and } Y = r_2)$$

For Independent Random Variables

- Theorem
  - Proof (cont'd)

$$E(XY) = \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1 \text{ and } Y = r_2)$$

$$= \sum_{r_1 \in X(S)} \sum_{r_2 \in Y(S)} r_1 r_2 \cdot P(X = r_1) \cdot P(Y = r_2)$$

$$= \sum_{r_1 \in X(S)} \left(r_1 \cdot P(X = r_1) \cdot \sum_{r_2 \in Y(S)} r_2 \cdot P(Y = r_2) \right)$$

$$= \sum_{r_1 \in X(S)} r_1 \cdot P(X = r_1) \cdot E(Y)$$

$$= E(Y) \left(\sum_{r_1 \in X(S)} r_1 \cdot P(X = r_1) \right) = E(Y)E(X)$$

Variance

- Deviation
  - The deviation of X at $s \in S$ is $X(s) - E(X)$, the difference between the value of X and the mean of X.

- Definition: Variance and Standard deviation
  - Let X be a random variable on the sample space S.
  - The variance of X, denoted by $V(X)$ is

$$V(X) = \sum_{s \in S} \big(X(s) - E(X)\big)^2 \cdot P(s)$$

  - That is $V(X)$ is the weighted average of the square of the deviation of X.
  - The standard deviation of X, denoted by $\sigma(X)$ is defined to be

$$\sigma(X) = \sqrt{V(X)}$$

Variance

- Theorem
  - If X is a random variable on a sample space S, then
$$V(X) = E(X^2) - E(X)^2$$
- Corollary
  - If X is a random variable on a sample space S and $E(X) = \mu$,
$$V(X) = E\big((X - \mu)^2\big)$$
- Example
  - For a random variable X,
    - $X(t) = 1$ if a Bernoulli trial t is a success,
    - $X(t) = 0$ if it is a failure.
  - p is the probability of success and q is that of failure.
  - What is the variance of the random variable X?

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p) = pq$$