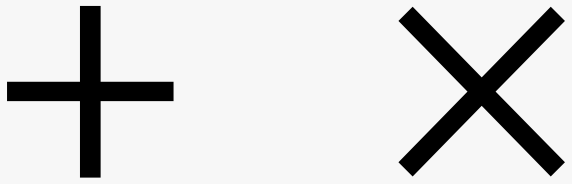
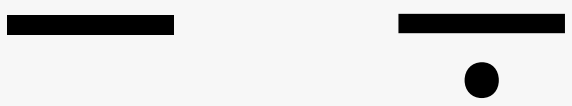
THEORY 1





2021

Contents

lesson checklist:

T1
T2
T3
T4

T5
T6
T7

next exam :
15/01/21

Sets

- a set is **unorderd** collection of distinct members (also called elements).

★ a set of positive integers bellow 10 ★

↳ **formal set-builder notation** = $S = \{ i \mid i \in \mathbb{I}, 0 < i < 10 \}$

↳ **Enumerating the set S** = $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

↳ any ordering is the same set

- shows that S contains values 'i', where 'i' is a member of the integers
  - The inequality shows it should be greater than 0 but less than 10.

enumeration = listing all values of a set

Cardinality = number of elements in a set

↳ #S, |S|, Card(S)

↳ $\emptyset$ = no elements

↳ # singleton set = 1

Finite set = A set whose elements can be numbered through from 1, 2 ... n, for some natural number n is a finite set.

↳ a set is finite when it contains a **bijection** (one to one correspondence), where $f: S \to \{1, \ldots n\}$

↳ sets that are not finite sets are **infinite sets**

Countable Sets

- all finite sets are countable sets
- can be finite or infinite
- if an infinite set is not countable it is uncountable

Sum + Product Rule

Sum Rule

- If there are n(A) ways to do A and, distinct from them, n(B) ways to do B, then **the number of ways to do A or B is n(A) + n(B)**. ... This is true if the number of ways of doing A and B are independent; the number of choices for doing B is the same regardless of which choice you made for A.

$|A+B| = |A| + |B|$

$n(A)+n(B)$

Addition Rule

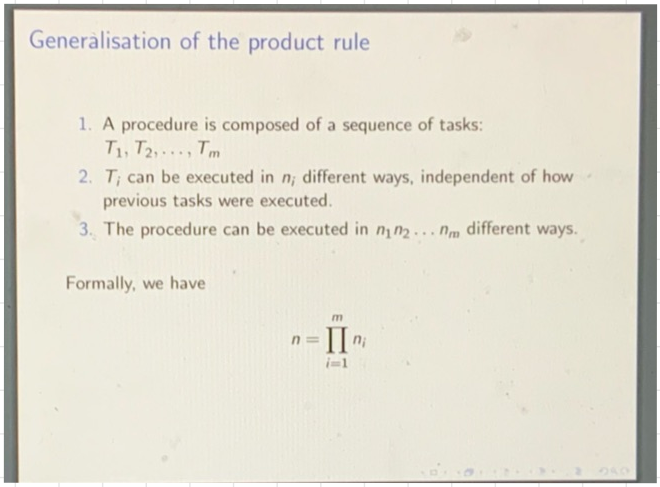
- The addition rule states **the probability of two events is the sum of the probability that either will happen minus the probability that both will happen**.

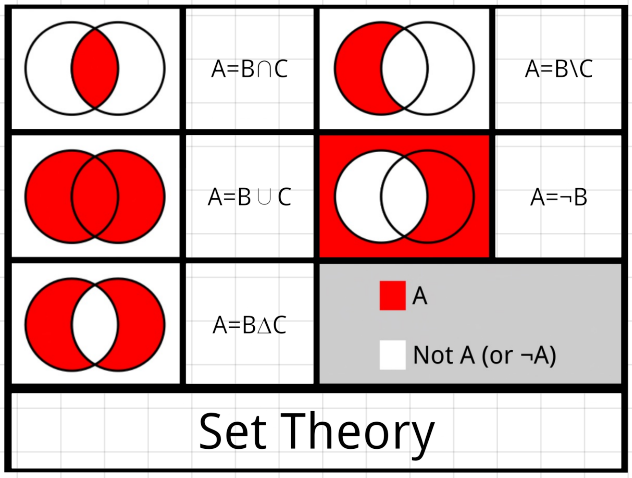
$P(A)+P(B)-P(A \cap B)$

Product Rule

- L If there are n(A) ways to do A and n(B) ways to do B, then the number of ways to do A and B is **n(A) × n(B)**. This is true if the number of ways of doing A and B are independent; the number of choices for doing B is the same regardless of which choice you made for A.

$n(A) \times n(B)$





$n = \prod_{i=1}^{m} n_i$

Subtraction + inclusion principle

Subtraction principle

- The cardinality of the union of two sets is the sum of the cardinality of The individual sets, with the cardinality of their intersection subtracted

$$\#(A_1 \cup A_2) = \#A_1 + \#A_2 - \#(A_1 \cap A_2)$$

inclusion / exclusion principle

The inclusion-exclusion principle

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| -$$

$$\ldots + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|$$

Here $|A|$ is the same as $\#(A)$, the cardinality of set A (used for more compact text formatting).

- The principle of inclusion and exclusion (PIE) is **a counting technique that computes the number of elements that satisfy at least one of several properties while guaranteeing that elements satisfying more than one property are not counted twic**

Division Rule + pigeonhole principle

Division Rule

- The division rule is a common way to ignore "unimportant" differences when you are counting things. You can count distinct objects, and then use the division rule to "merge" the ones that are not significantly different.
  We will state the Division Rule twice, once informally and then again with more precise notation.
  Theorem 1.1 (Division Rule). If B is a finite set and f : A → B maps precisely k items of A to every item of B, then A has k times as many items as B.
  For example, suppose A is a set of students, B is a set of tutorials, and f defines the assignment of students to tutorials. If 12 students are assigned to every tutorial, then the Division Rule says that there are 12 times as many students as tutorials.

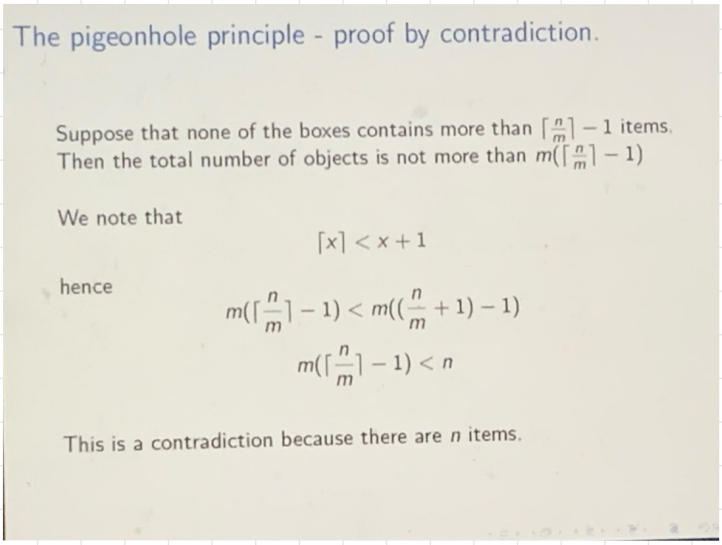
Floor function: Rounds down to nearest int: $y = \lfloor x \rfloor$

ceiling function: Rounds up to the nearest int: $y = \lceil x \rceil$

Round to the nearest int. expressed as : $y = \lceil x \rfloor$

Pigeonhole principle

- let n and m be positive integers
  • If n items are pwh into m boxes where n> m,
  then at least one box must contain more than one item
  • if n items are placed into m boxes, then there is
  at least one box containing-at least $\lceil (n/m) \rceil$ objects



The pigeonhole principle - proof by contradiction.

Suppose that none of the boxes contains more than $\lceil \frac{n}{m} \rceil - 1$ items.
Then the total number of objects is not more than $m(\lceil \frac{n}{m} \rceil - 1)$

We note that

$$\lceil x \rceil < x + 1$$

hence

$$m(\lceil \frac{n}{m} \rceil - 1) < m((\frac{n}{m} + 1) - 1)$$

$$m(\lceil \frac{n}{m} \rceil - 1) < n$$

This is a contradiction because there are n items.

Combinations + permutations

Permutations
- a set of arrangements constructed from a set of elements, where the order of the elements is relevant.

Combinations
- order of elements in a set is irrelevant

An ordered arrangement of r elements from a set of n objects is called an r-permutation.
An unordered arrangement of r elements from a set of n objects called an r-combination.

- When repetitions are not allowed: $1 \le r \le n$.

Suppose that our set of distinct objects is the set of 10 digits: {0, 1, . . . , 9} and we compose 3-permutations.

1. With allowed repetition of elements : 333, 343, 345, 456 are all valid.
   Example: all four of these numbers could be used on a combination lock (which should be called a permutation lock!)

2. Without allowed repetition of elements : 333, 343 are invalid, 345, 456 are valid.
   Example: the latter two numbers could be the numbers of the first three athletes in a race, but the first two could not.

r- permutation without repition

↳ $P(n,r) = \frac{n!}{(n-r)!}$

- when we chose r components, they can be ordered in r! ways.

In a combination, all those orderings are the same thing, so we need to cancel out a factor of r! by using the division rule, which gives us a formula for an r-combination as:

↳ $C(n,r) = \frac{n!}{r!(n-r)!}$

Combinations without repetition: the binomial coefficient

In general, for combinations that exclude repetition, we use the binomial equation (informally, we say *"n choose r"*):

$$C(n,r) = {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}, \quad n \geq r \qquad (3)$$

- In our earlier example, we chose 3 decimal digits from the 10 available. Here we have $\frac{720}{3!} = 120$ combinations without repetition.
- In general ${}_nC_r$ is the number of subsets with r elements selected from a set with n elements
- The binomial coefficient gives the number of subsets with r elements for some set of cardinality n.

Symetry of the binomial Coefficient

↳ $nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{n-r}$

? So choosing r from n has the same number of combinations as choosing (n – r) from n

Combinations with repetition

1. Suppose that we arrange the set of decimal digits 0 . . . 9 in some arbitrary order and we imagine a pointer pointing to the first digit.
2. We then imagine a binary string that has the following interpretation:
   - 1 means that we should select that digit, multiple 1s means select that digit multiple times (up to 3 times in our '10 choose 3' example)
   - 0 means that we should move the pointer to the next digit in the sequence of 10. There should be 10-1=9 moves.
3. Therefore, in our '10 choose 3' example, we need three 1s in our binary string and nine zeros.
4. So our specific problem of *finding combinations of 3 decimal digits with repetitions allowed* is mapped to *finding all 12-bit binary strings that have exactly three 1s.*

Table: Permutation and combination formulae.

	Repetitions	No repetitions
Permutations	n^r	$\frac{n!}{(n-r)!}$
Combinations	$\frac{(r+n-1)!}{r!(n-1)!}$	$\frac{n!}{r!(n-r)!}$

↳ ? Count the combinations of choosing r from n with repetitions allowed.
transforms to
? Count the combinations of choosing r from r + n – 1 with repetitions NOT allowed.
In the transformed problem, we can use the binomial coefficient formula again, but substitute r + n – 1 for n and simplify:

↳ $${}_{(r+n-1)}C_r = \binom{r+n-1}{r} = \frac{(r+n-1)!}{r!(n-1)!}$$

Binomial Coefficients + Derangements

- A Binomial Coefficient is of the form $(x+y)^n$.
- its Binomial because there is a sum of two terms

- For example:
  $(x+y)4 = (x+y)(x+y)(x+y)(x+y) = x4+4x3y+6x2y2+4xy3+y4$

↳ Notice that the sequence of coefficients in this equation is 1-4-6-4-1. This is palindromic and is related to the symmetry in the binomial coefficient that we discussed in the previous mini-lecture.

- Select r of the y terms and (n–r) of the x terms from the four (x + y) terms and multiply them together. We need to cover all of the r-combinations from n products. This is where 'n choose r' binomial coefficient comes in. We do this for all possible values of r and add.
  So in general

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

For our case where n = 4

? For the y4 term, there is only one combination that achieves this, we must choose a y from each of the (x + y) terms.
? Similarly, there is only one way to form the product x4, we must choose an x from each of the (x + y) terms.
? There are four ways to form the term x3y (similarly for xy3).
? For the x2y2 term, we choose two y terms from the four (x + y) terms (with the remaining two defaulting to x) and this can be done in '4 choose 2', which is six ways.

Suppose we place the coefficients of $(x+y)^n$ on the nth row and let the columns in each row of the triangle run from $r=0$ to $r=n$.

$n=0$						1						
$n=1$					1		1					
$n=2$				1		2		1				
$n=3$			1		3		3		1			
$n=4$		1		4		6		4		1		
$n=5$	1		5		10		10		5		1	
$n=6$	1		6		15		20		15		6	1

Pascal's identity:

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}, \quad r \le n \qquad (1)$$

- Pascal's identity can be understood by reasoning about subsets of r-combinations. Suppose we select an arbitrary element, X, from set S of cardinality n.

  - ? The term coloured in blue counts all subsets that contain X.
  - ? The term coloured in green counts all subsets that don't contain X.
  - ? All r-combinations of S either contain X or do not contain X.
  - ? The two subsets above are disjoint, meaning we can apply the sum rule.
  - ? Hence Pascal's identity is proved

- A derangement is a permutation of all of the elements of a finite set, that leaves no object in its original position.

The number of derangements of a set with n elements is described by the subfactorial of n, with symbol !n, where:

$$!n = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right]$$

The term in the square brackets is the Taylor Series expansion of ex where e = 2.718 . . . is Euler's number and x = –1.

Hence:

$$!n \approx \frac{n!}{e}$$

rounding this quotient to the nearest integer

$$!n = \left[\frac{n!}{e}\right]$$

Discrete Probability

Uncertainty

- Logic deals with things that are true or False

1. Probability theories deal with situations or experiments whose outcomes cannot be predicted with certainty.
2. Hence such theories allow us to reason about events in the presence of uncertainty.

Discrete + Continuous probability

1. In the case of discrete probability, the outcome is a member of a defined set of values.

2. In contrast continuous probability deals with outcomes that can have a continuous set of values: 1.617.., 1.735.., 1.821.., 1.919..

3. Usually probability theories can be applied to both outcome types and have analagous forms.

Finite discrete Sample Spaces

- set of all possible outcomes in an experiment
- Discrete probability distributes over these outcomes

probability of events + Combined events

- independent events
- Disjoint events

Sample Spaces + Outcomes

- The first assuption of probability theory is that there is a known samplespace

- The sample space, Ω, is the set of all possible outcomes of an experiment.

Sample Space for 3 Coin flips

Ω = {H,T}3 = {TTT,TTH,THT,THH,HTT,HTH,HHT,HHH}
? Each member of the sample space is called an outcome, ω, or an elementary event, so Ω = {ω1 . . . ω8}.
? Typically we are interested in a (possibly more complex) event that is a subset of the sample space.
? Events in our example could be
? event EA occurs if the outcome is in the set of outcomes where there is exactly one head: A = {TTH , THT , HTT }
? event EB occurs if the outcome is in the set of outcomes where there are three heads: B = {HHH}
? Event EB is an elementary event, whereas event EA is not.

Notation

The definition of the probability of an event can be shown as:

$$P(E_A) = P(\omega \mid \omega \in A, A \subseteq \Omega)$$

This reads as: the probability of event A is the probability of outcome omega, such that omega is a member of the set of outcomes A, where set A is a subset of the sample space.
To simplify notation, we may refer to the probability of event A, EA, more straightforwardly as P(A). In other words the argument of P(.) may be any subset of Ω, which includes Ω itself and the empty set ∅ or any complex set expression involving intersections and/or unions of various sets

Outcomes for three flips of a coin

- ? Ω = {TTT,TTH,THT,THH,HTT,HTH,HHT,HHH}
- ? A: the set of outcomes where there is exactly one head.
- ? $\bar{A}$: the set of outcomes where there is NOT exactly one head. ? NotethatP(A∪$\bar{A}$)=P(Ω)=P(A)+P($\bar{A}$)=1.
- ? NotethatP(A∩$\bar{A}$)=P(∅)=0

The second assumption of probability is that every outcome, ω, of a sample space is assigned some probability, P(ω), where P is a real function of ω.
? We can form an event A as any set of outcomes that is a subset of Ω, i.e. A ⊆ Ω.
? The probability of an event A is defined as: P(A) = ? P(ω) ω∈A
? The probability of this event is: 0 ≤ P(A) ≤ 1.
? 0 indicates that the event is impossible and 1 indicates that it is certain. In particular:
? P(∅)=0 ? P(Ω)=1

Finite discrete probability space

A finite discrete probability space, (Ω,P), is a finite set of outcomes ω, called a sample space, Ω, together with a function
P : Ω → R, that maps the set of outcomes to a real number. This function is called a probability distribution that satisfies the following properties:

- $0 \leq P(\omega) \leq 1, \forall \omega \in \Omega$
- $\sum_{\omega \in \Omega} P(\omega) = 1$

[?] The function P is called a probability distribution on Ω, as it distributes the probability of 1 over all of the outcomes ω ∈ Ω.
[?] The probability of an event is the sum of the probabilities of the set of outcomes that define that event:

$$P(A) = \sum_{\omega \in A} P(\omega)$$

[?] A probability distribution may be called a probability mass function in some texts on discrete probability.

Probability of Combined events

[?] The probability of event A AND event B occurring: $P(E_A \wedge E_B) = P(A \cap B)$

[?] The probability of event A OR event B occurring: $P(E_A \vee E_B) = P(A \cup B)$

[?] The probability of event A NOT occurring: $P(\neg E_A) = P(\bar{A})$

Probability of (A ∩ B) for independent events

- Given any finite discrete probability space (Ω,P), then for any two independent events A, B ⊆ Ω: [?] P(A ∩ B) = P(A)P(B)
  [?] Two events are dependent if they are not independent.

Probability of (A OR B) for disjoint events

- Disjoint events cannot happen at the same time i.e. they are mutually exclusive. Given any finite discrete probability space (Ω, P), then for any two disjoint events A, B ⊆ Ω
- P(A ∪ B) = P(A) + P(B) ThisrequiresthatP(A∩B)=0 ⟹ A∩B=∅

General formulation for P(A ∪ B)

- Given any finite discrete probability space (Ω,P), then for any two events A, B ⊆ Ω:
  P(A ∪ B) = P(A) + P(B) – P(A ∩ B)
- This is related to the subtraction rule, with events described by more complex set unions relating to the inclusion-exclusion principle.
  Finally, we also note the following [?] $P(A^-) = 1 - P(A)$
  [?] A⊆B ⟹ P(A)≤P(B)

Conditional probability + bayes Rule

- In general, the occurrence of some event A changes the probability that another event B occurs. So we can improve our calculation of the probability of event B happening - if we take into account the prior knowledge that event A has occurred

Bayes Rule

$$P(B|A) = P(B \cap A)/P(A), \ P(A) \neq 0$$

$$P(A|B) = P(A \cap B)/P(B), \ P(B) \neq 0$$

Note that set intersection is commutative

$$B \cap A = A \cap B$$

hence

$$P(B|A)P(A) = P(A|B)P(B)$$

Thus we have Bayes' rule

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \ P(A) \neq 0$$

- Bayes' rule provides us with a relationship between the conditional probabilities of a pair of events. This turns out to be extremely useful.

Expand the denominator to express $P(A)$ in terms of conditional probabilities

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

Often event B describes some state: eg - someone either has a disease or doesn't. Event A would then describe the outcome of a related measurement: eg - high blood pressure - i.e. higher than some critical threshold.

- P(B) is the prior probability of the state B, which is the proportion of people that have the disease in the population. This is our answer if we have no further evidence.
- P(AIB) is the liklihood function giving the probability of a positive test for high blood pressure, state A, given the state B.
- P(BIA) is the a posteriori probability of the state B given the test A. So this is a more informed, and therefore more useful estimate of the probability of B than the prior.

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

Independence - revisited

Any two events A, B, where P(A)/= 0 and P(B)/= 0 are independent when any of the following three statements hold.

- P(A ∩ B) = P(A)P(B)
- P(BIA) = P(B)
- P(AIB) = P(A)

These statements are equivalent, by inspection of Bayes' rule, given again below

Discrete Random Variables

- maps the outcomes of a random process to numerical values

Random variables can be used to describe events very compactly. For example, in the three coin flip experiment, consider the probabilities of the following events:
[?] P(Number of heads is exactly two) [?] P(Number of heads is at least two)
If the random variable X represents the number of heads in this experiment, we have the more compact event statements:
[?] P(X = 2) [?] P(X ≥ 2)

Uniform distributions

In many cases, the probability distribution is uniform. This means that the probability is the same for all ω ∈ Ω.
[?] For a flip of a fair coin: P(H) = P(T) = 12.
[?] For a roll of a fair die: P(1) = P(2)⋯ = P(6) = 16.
In uniform distributions, computation of probabilities is straightforward. P(A) = nA where nA is the number of outcomes
N
in the event and N is the number of outcomes in the sample space, Ω.

Binomial distribution

The binomial distribution is the distribution of the number of successes r in n experiments each of which have binary outcome, with an outcome of 1 occurring with probability p, where 0 ≤ p ≤ 1 and outcome of 0 occurring with probability q = 1 − p.

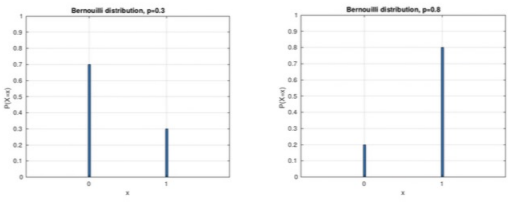
$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}, \quad \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Cumulative probability distribution

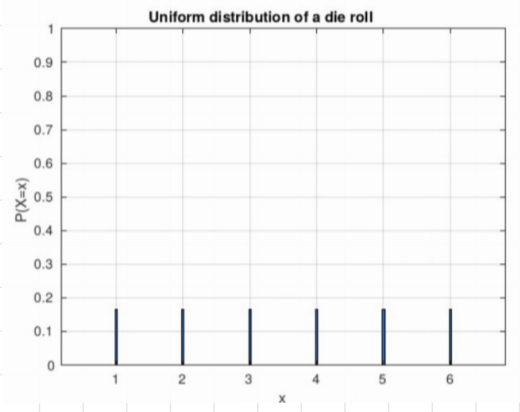
Often we want to know the probability that a random variable is less than or equal to some value.
[?] This is called the cumulative probability.
[?] "Cumulative", because we need to sum the probabilities for all values that are less than or equal to the specified value.

Bernouilli distribution

A *Bernouilli distribution* is the distribution of a random variable that has a binary outcome, $\Omega = \{0, 1\}$. It takes on value 1 with probability p, where $0 \leq p \leq 1$ and it takes on value 0 with probability $q = 1 - p$. Examples:



Note that a flip of a coin has a Bernouilli distribution with $p = 0.5$.



Discrete probability

Expectation + Variance

- The probability distribution gives comprehensive information about a random variable.
- However, we often like to summarise the probability distribution of a random variable.
- This brings us to the concept of expectation or expected value - a kind of weighted mean.
- We can also compute the variance of a random variable.
- In general, these values are not integers

expectation of a discrete random variable

- Given a finite discrete probability space (Ω,P), for some random variable X, the mean value or expected value or expectation of X is the number E[X] such that

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

- It's a weighted mean, where the weights sum to unity, because the weights are probabilities.
- Often denoted by the symbol μ or μX .

properties of expectation

- Given any two random variables, X,Y on a discrete probability space, we have the linearity property (holds even if X and Y are not independent):
  - E[X+Y]=E[X]+E[Y]
  - E[αX + β] = αE[X] + β, for real valued constants α and β

  Note that a product of two random variables is a non-linear relation but, for independent random variables, X , Y :
  - E[XY] = E[X]E[Y]

Expected value of three important distributions.

We can determine general formulae for the expected value of various distributions.

- Expected value of a Bernouilli distribution

E[X] = (0 × (1 – p)) + (1 × p) = p

- For a binomial distribution, X is now the sum of n Bernouilli trials, hence by the linearity of expectation, we have:

E[X] = np.

- Uniform distribution: E [X] is the mean of X (ω) over all ω ∈ Ω

$$Var(X) = E[(X - E[X])^2]$$

$$Var(X) = \sum_{\omega \in \Omega} (X(\omega) - E[X])^2 P(\omega)$$

Variance of direct random variables

- The variance of a random variable measures the degree to which the values of a random variable differ from the expected value.
- Computed as the expectation of the square difference $(x - E[x])^2$

Denoted as : $var(x)$ or σ^2 or σ^2_x

Alternative expression for variance

Note the linearity of expectation and the fact that the expectation of a constant is itself. We use these in the third line of the analysis below

$$\begin{aligned} Var(x) &= E[(X - E[x])^2] \\ &= E[x^2 - 2xE[x] + (E[x])^2] \\ &= E[x^2] - (E[x])^2 \end{aligned}$$

Joint probability distributions

Consider a pair of possibly related random variables (r.v.) X and Y . We now consider:
[?] Joint probability distributions.
[?] Marginal probability distributions.
[?] Conditional probability distributions. [?] Expectation and covariance.
[?] Correlation and independence.

Joint histogram example

Suppose you flip a coin and roll a die 60 times and record the results. A typical tabulated result could be:

		Die						
		1	2	3	4	5	6	
Coin	0	6	3	5	5	3	6	28
	1	4	8	7	3	7	3	32
		10	11	12	8	10	9	60

- The tabulated numbers in black is a joint histogram.
- The tabulated numbers in red are marginal histograms.

Joint distribution example

Suppose $N_{trials} \to \infty$, then we converge towards the joint probability distribution:

		Die						
		1	2	3	4	5	6	
Coin	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
	1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
		$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

- The tabulated numbers in black is the joint distribution.
- The tabulated numbers in red are marginal distributions.
- Marginal distributions are for some prescribed subset of the full set of random variables. They ignore other r.v.s by 'summing them out' - which is also known as marginalising them - they are said to be 'marginalised out'.
- Note that the marginal distributions for both X and Y sum to unity.

We can compute various joint probabilities, joint probability distributions and statistics of joint probability distributions.
[?] The probability that X is x and Y is y: a joint probability [?] P(X =x,Y =y)
[?] The probability distribution of Y given some X: a conditional probability
[?] P(Y|X=x)
[?] The expectation of Y given some X: a conditional expectation
[?] E[Y|X=x]

Covariance

Covariance of two r.v.'s:
[?] Covariance is a measure of the joint variability of two random variables.
[?] It is the expected value of the product of the deviation of X from its expected value and the deviation of Y from its expected value.

$$Cov(X,Y) = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j - \mu_X)(y_i - \mu_Y)P(X = x_j, Y = y_i)$$

Note:

$$Cov(X,Y) = Cov(Y,X)$$

and

$$Var(X) = Cov(X,X)$$

Interpreting Covariance

Consider the term:
$(x_j - \mu_X)(y_i - \mu_Y)$
When X is above its mean, is Y usually above its mean? If so:
- When X is below its mean then Y will usually be below its mean.
- In general as one r.v. increases/decreases the other increases/decreases
- In this case, the term above will usually be positive - and, aggregated over all random vectors in the distribution, will sum to a positive value

Correlation

- Covariance can be normalised such that it is bounded between −1 and +1, and this measure is called correlation.

$$Corr(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

This indicates the strength of the linear relation between the two random variables, and hence is easier to interpret than unnormalised covariance.

Covariance + independence

- Random variables with zero covariance are uncorrelated.
- If two random variables are independent, then their covariance is zero.
- If two random variables are uncorrelated, this does not imply their independence.
- Example: let X be uniformly distributed in [−1, 1] and let Y =X2.
- The relationship implies that X and Y are not independent. However, their covariance can be shown to be zero.

Discrete probability summary

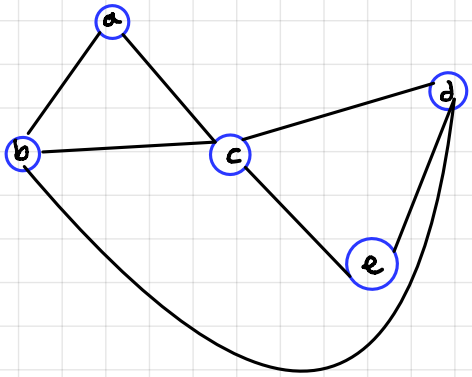
After studying Theory-1 lectures on discrete probability you should be able to explain and use the following concepts in discrete probability problem solving:

1. Discrete sample spaces and the probability of events
2. Conditional probability and Bayes' rule
3. Discrete random variables and their probability distributions
4. Statistics of random variables: expectation and variance
5. Discrete random vectors: joint probability distributions, covariance, correlation and independence.

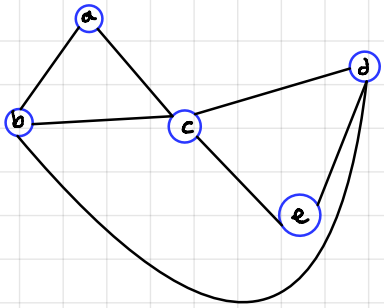
Graphs

• a finite graph is represented by an ordered pair of finite sets G = (V, E)

↳ the objects are represented by a finite set of vertices in the graph. (blue circles)
↳ the relations between objects are represented by a finite set of edges in the graph:(black lines)
↳ each edge in the set E must start at a vertex and end at a vertex



Networks
• Networks + graphs are used interchangeably
• networks use terms like nodes + links whereas graphs use vertices + edges
↳ a graph whose s vehicles are indistinguishable is more likely to be called a network.
networks = more physical, she has routers and computers
graphs = More abstract mathematical structures



$G = (V, E)$

$V = \{a, b, c, d, e\}$

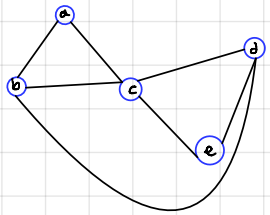
$E = \{ab, ac, bc, bd, cd, ce, de\}$

What relation do the graph edges represent? :

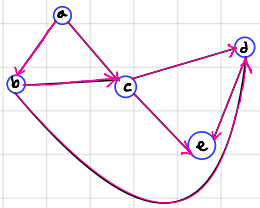
$$1 \leq |f(v_i) - f(v_j)| \leq 2$$

- The order of a graph is the cardinality of the vertex set #V
- The size of a graph is the cardinality of the edge set #E

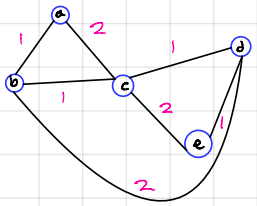
Types of Graphs



• undirected
• unordered pairs of vertices

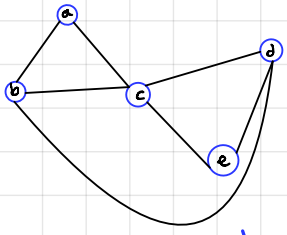


• directed
• edges are ordered pairs
• digraph

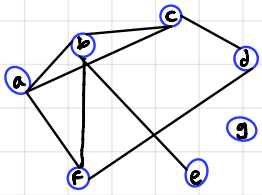


• weighted
• all edges have a numerical weight
• graphs can be weighted and directed

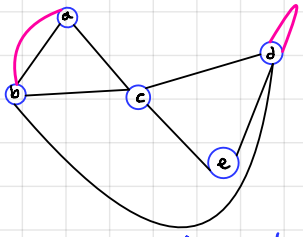
Graphs



simple, connected



simple, disconnected



multigraph

Simple graph: undirected, unweighted, no loops, no multiple edges

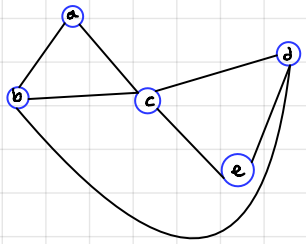
Degrees

- The degree $\deg(v)$, of a vertex, v, is the amount of edges attached to that vertex
- The degree sequence of an undirected graph is the non-increasing sequence of its vertex degrees

Min graph degree = $\delta(G)$

Max graph degree = $\Delta(G)$

- A digraph has an in degree, $\deg^+(v)$, and an out degree, $\deg^-(v)$, representing the number of edges entering and leaving the vertex.



- the number of odd-degree vertices is always even.

$\deg(a) = 2$

$\deg(b) = 3$

$\deg(c) = 4$

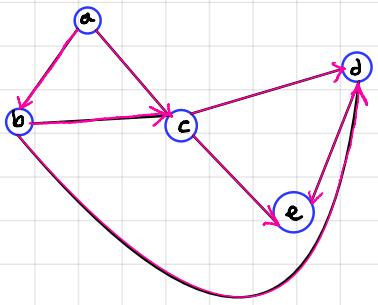
$\deg(d) = 3$

$\deg(e) = 2$

degree sequence = $(4, 3, 3, 2, 2)$

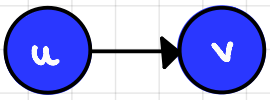
$$\sum_{v \in V} \deg v = 2\#E$$

$$\sum_{v \in V} \deg^-(v) = \sum_{v \in V} \deg^+(v) = \#E$$



	a	b	c	d	e	Σ
$\deg^-$	0	1	2	2	2	7
$\deg^+$	2	2	2	1	0	7
$\deg$	2	3	4	3	2	14

Incidence



• suppose that (U, V) is on edge in a graph G = (v, e) there is an incidence relation between the edge + its associated vertices

• if (U,V) is an edge in an undirected graph G=(V,E) then (U,V)is an incident to vertices u + v

if (u,v)is an edge in an undirected graph G= (V, E) then (U,V) is incident from vertex u and is incident to vertex v

• when the graph is undirected the adjacency relation is symmetric, two vertices are mutually adjacent
• in a directed graph vertex v is adjacent to a, if there is an edge that leaves u and enters v.

Graph with large number of nodes but relatively few edges has a sparse adjacency matrix. An alternative is to use an adjacency list.

vertex	adjacencies
a	b, c
b	a, c, d
c	a, b, d, e
d	b, c ,e
e	c, d

Adjacency matrix

• used to represent a finite graph
• the elements of the matrix indicate wether pairs of Vertices are adjacent or not

Adjacency matrix example, directed graph

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- The directed relation goes from a vertex in a *row* of the matrix to a vertex in the *column*.
- Out-degree is the sum of rows, in-degree is the sum of columns.
- The adjacency matrix for an directed graph is not necessarily symmetric.

Adjacency matrix example, undirected graph

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

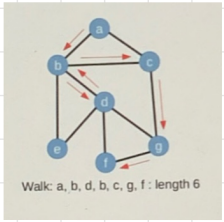
- Vertex degree is the sum of the relevant row (or column).
- The adjacency matrix for an *undirected* graph is symmetric.

Walks, Trails, Paths

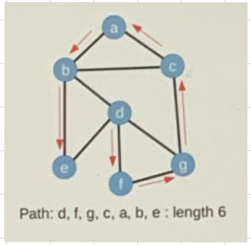
• a walk in a graph is a sequence of vertices $v_1, v_2 \dots v_k$ such that $(v_i, v_{i+1}) \in E$, $i = 1, 2, \dots k-1$.

• If the edges in a walk are distinct, then the walk is called a trail.
• If the vertices V1+V2…Vk are distinct, then the walk is called a path.
trails C walks; paths C trails
• the length of a walk, trail or path is its number of edges

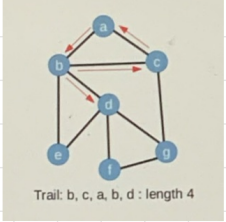
not distinct vertices → walk
distinct vortices → path



Walk

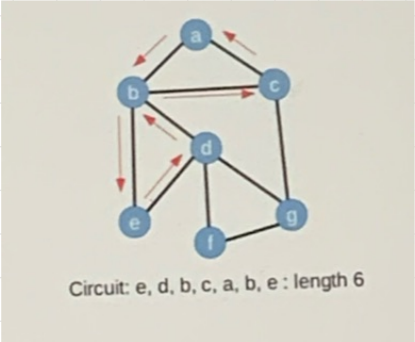


Path



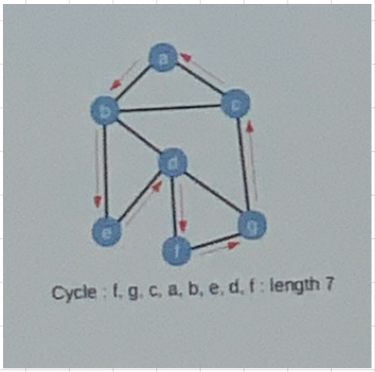
trail

Distance Between Vertices + graph diameter



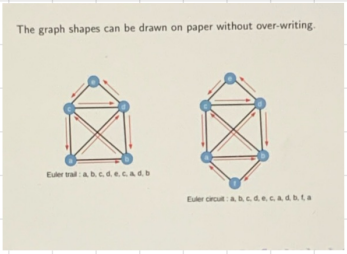
• the distance between two vertices (U,V)in a graph is the number of edges in a shortest path Connecting them. It is also called a graph geo disc. There may be more than one shortest path between two vertices.

• A distance matrix, D, is a square matrix contains the pairwise distances, d u v, between vertices.
• If one vertex cannot be reached from another by a path made of edges, then we say their distance from each other is infinity . this means the graph is disconnected .
• the diameter of a graph is the Max distance over all pairs of vertices



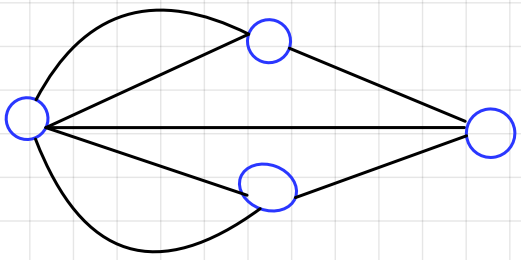
Circuit + Cycles

• a trail that begins and ends at the some vertex is called a closed trail, or a circuit, a closed path is called a cycle, a path v1, v2...Vk together with the edge Vk, v1 and with K≥3
the length of a circuit or cycle is its number of edges.



Euler on the Seven bridges

• the chosen start + end point land masses must have an odd number of bridges (edges)
• all other vertices must have an even number of edges
• but all nodes have an odd number of edges
> no solution <



Eulerian trails, circuits and graphs

Eulerian = visits each edge exactly once
> can exist if the start + of vertices have an odd
degree and all others have an even degree
Eulerian trail = a trail that crosses every edge once
Eulerian circuit = a trail that ends at the initial Vertex

Classical problems

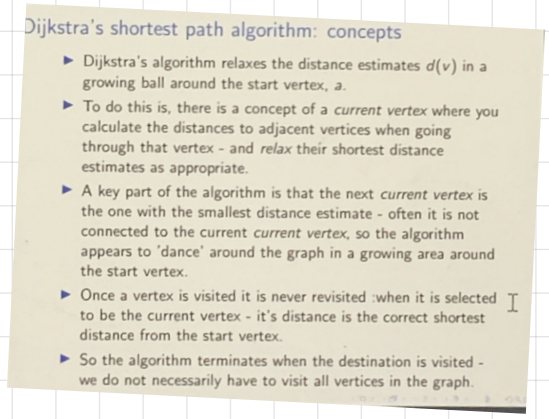
Traveling salesman:
• suppose we have N Cities and we know the travel distance between all pairs of cities.
> from some starting city, what path should we take seen that we visit each city once and end back at the starting city, in the minimum distance?
> so this problem is to find an optimal Hamiltonian cycle on a weighted graph, where the weights are the distances (or times) between cities. The graph could be directed [i. e. one way streets] or undirected.

Shortest path problem

• finds a path between two vertices in a weighted graph such that the sum of the weights of the paths edges is minimised.
Floyd-Wortham - finds shortest paths between all vertex pairs
Dijkstra's - find shortest path between a closer starting vertex and another.

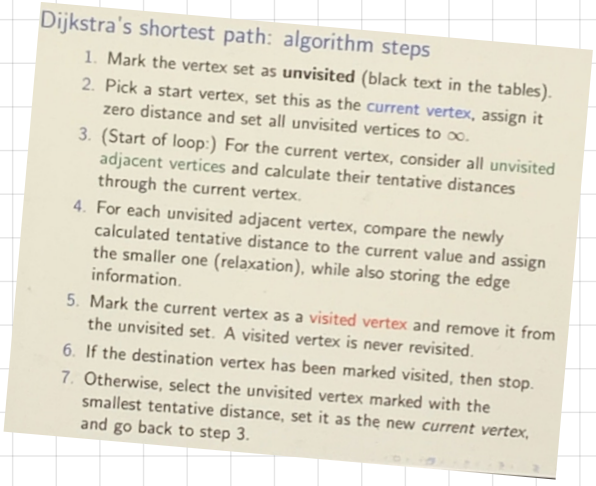
Dijkstra's shortest path - iteration

• Iterative procedures are used to refine estimates
• start with an initial estimate of distance of all vertices from the start vertex a. Start vertex itself is 0. d(a,a)=0. all other are infinity.
• Iteratively refine such upper bands for d(a,v).the distance of some vertex V From the start of vertex a. The refinement step is called relaxation.



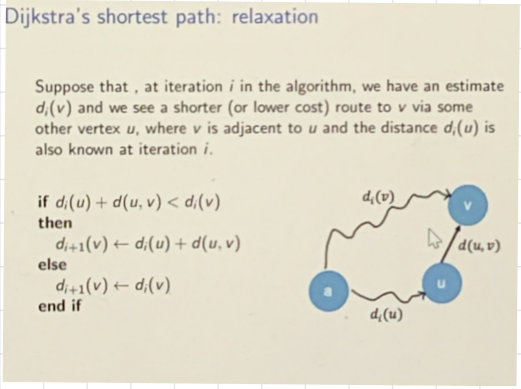
Dijkstra's shortest path algorithm: concepts

- Dijkstra's algorithm relaxes the distance estimates $d(v)$ in a growing ball around the start vertex, a.
- To do this is, there is a concept of a *current vertex* where you calculate the distances to adjacent vertices when going through that vertex - and *relax* their shortest distance estimates as appropriate.
- A key part of the algorithm is that the next *current vertex* is the one with the smallest distance estimate - often it is not connected to the current *current vertex*, so the algorithm appears to 'dance' around the graph in a growing area around the start vertex.
- Once a vertex is visited it is never revisited :when it is selected to be the current vertex - it's distance is the correct shortest distance from the start vertex.
- So the algorithm terminates when the destination is visited - we do not necessarily have to visit all vertices in the graph.



Dijkstra's shortest path: algorithm steps

1. Mark the vertex set as **unvisited** (black text in the tables).
2. Pick a start vertex, set this as the current vertex, assign it zero distance and set all unvisited vertices to ∞.
3. (Start of loop:) For the current vertex, consider all unvisited adjacent vertices and calculate their tentative distances through the current vertex.
4. For each unvisited adjacent vertex, compare the newly calculated tentative distance to the current value and assign the smaller one (relaxation), while also storing the edge information.
5. Mark the current vertex as a visited vertex and remove it from the unvisited set. A visited vertex is never revisited.
6. If the destination vertex has been marked visited, then stop.
7. Otherwise, select the unvisited vertex marked with the smallest tentative distance, set it as the new *current vertex*, and go back to step 3.



Dijkstra's shortest path: relaxation

Suppose that , at iteration i in the algorithm, we have an estimate $d_i(v)$ and we see a shorter (or lower cost) route to v via some other vertex u, where v is adjacent to u and the distance $d_i(u)$ is also known at iteration i.

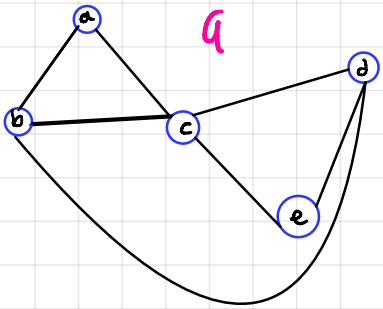
if $d_i(u) + d(u,v) < d_i(v)$
then
  $d_{i+1}(v) \leftarrow d_i(u) + d(u,v)$
else
  $d_{i+1}(v) \leftarrow d_i(v)$
end if

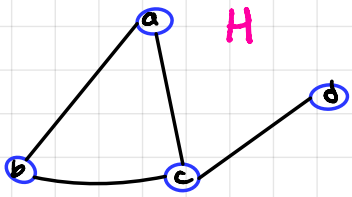
Subgraphs

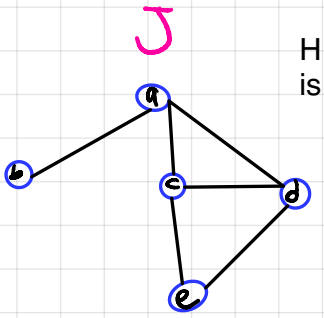
• A Subgraph is a graph who's vertices + edges are subsets of the vertices + edges of another graph
Two graphs:

$G = (V(G), E(G))$ and $H = (V(H), E(H))$

$H \subseteq G \leftrightarrow V(H) \subseteq V(G) \wedge E(H) \subseteq E(G)$



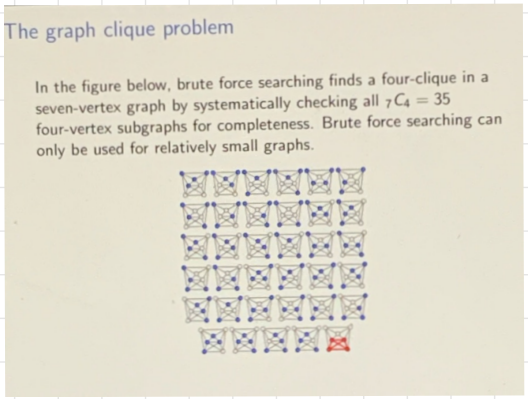




H is a subgraph of G but J is not a subgraph of G

The graph clique problem

• a clique is a subgraph where all vertices are adjacent
↳ also called a Complete subgraph
↳ maximal clique problem is to find the clique with the largest number of vertices

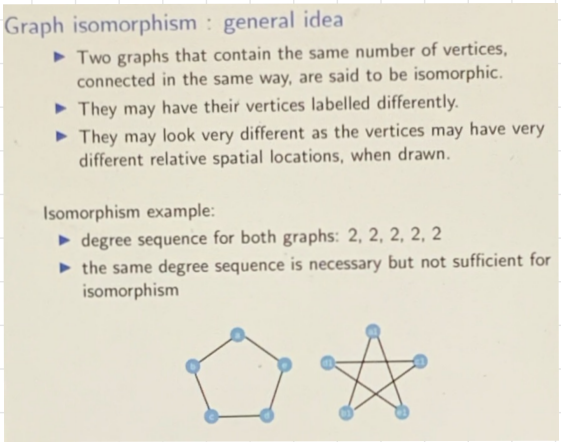


isomorphism

• graphs a + H are isomorphic if there is a bijection between the vertex sets of G and H

$f: V(G) \to V(H)$

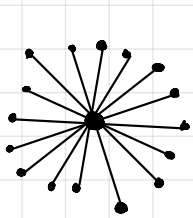
• graph isomorphism has a structure preserving bijection.

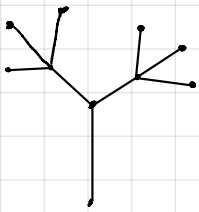


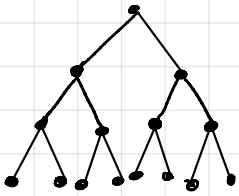
Graphs

Trees

- an undirected acyclic graph
- any vertex pair is connected by a single path

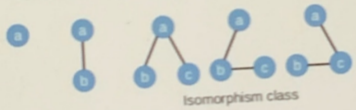






For a known graph order, $n = \#V$, how many trees are there?

- For the case of $n = 1$ and $n = 2$ we have one tree, and for $n = 3$, we have three trees, but they are an isomorphism class.



In general, Cayley's formula holds for the number of trees, N_T:

$$N_T = n^{n-2}$$

However, this counts all trees within isomorphisms. The number of Non-Isomorphic Trees (NIT) is smaller.

n	1	2	3	4	5	6	7
#NIT	1	1	1	2	3	6	11

For low order graphs, these can be found by systematically considering the trees with the highest possible maximum degree $(n-1)$, which is a star, to the lowest possible maximum degree (2), which is a chain.

Depth first search = Starts at the root vertex and explores as far as possible along each branch before backtracking
breadth first search = Searches all vertices at current depth
forest = undirected acylic graph
↳ trees C forests
leaf = a vertex of degree one on a tree

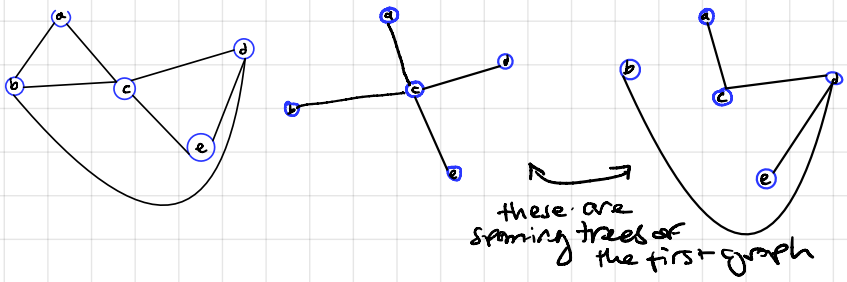
Relationship between size + order

let $n = \#v$, $m = \#$, Tree = T

$m = n - 1$

Rooted + binary trees

Rooted trees = has a special vertex known to be the starting vertex called the Root.
Binary trees = has at most two descending branches at any one vertex. It can be both rooted + binary.
Spanning trees - it is a tree that is a subgraph of G that includes every vertex

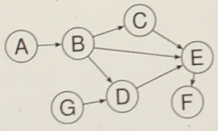


Minimum spinning tree - a spanning tree whose sum of edge weights is as small as possible

Directed acyclic graphs

Directed Acyclic Graphs (DAGs)

- A DAG has a topological ordering: every edge is directed from earlier to later in the sequence.
- A vertex is defined as a *source* vertex if all of its edges are incident from it.
- A vertex is defined as a *sink* vertex if all of its edges are incident to it.
- Every DAG has at least one source vertex and at least one sink vertex.
- In our example we have two source vertices (A, G) and one sink vertex (F).



DAG applications

Generally, suitable for modelling any system that has strictly feedforward data flows (i.e. no feedback allowed).

- Computer instruction scheduling : scheduler has to handle dependencies correctly.
- Spreadsheets : topological ordering of a DAG can be used to update all dependent cell values when some spreadsheet cell entry is changed.
- Combinational logic, in electronic circuit design.
- Feed forward neural networks: deep learning is a highly active area of research.
- Family 'trees' (inverted commas, as they are not really trees).

Truth Values

- A statement is a collection of symbols that has a truth value – either False (F) or True (T)
  - ↳ • is Paris in France? = T
  - • 2+2 = 6 = F
- Symbols called Connectives are used to form larger statements out of smaller ones:

$\wedge$	and
$\vee$	or
$\neg$	not

- A Conjuction is a compound statement in which two substatements are connected by $\wedge$
- A disjunction is a compound statement in which two substatements are connected by $\vee$
- The negation of a statement p is $\neg p$. Intuitively 'p is not the case.'
- A proposition is a statement in which the basic substatements are variable, each with F or T as possible values.

Conjunction

p	q	$p \wedge q$
F	F	F
F	T	F
T	F	F
T	T	T

disjunction

p	q	$p \vee q$
F	F	F
F	T	T
T	F	T
T	T	T

negation

p	$\neg p$
F	T
T	F

- Here is a truth table for the proposition $\neg(p \wedge \neg q)$:

p	q	$\neg q$	$p \wedge \neg q$	$\neg(p \wedge \neg q)$
F	F	T	F	T
F	T	F	F	T
T	F	T	T	F
T	T	F	F	T

Propositions

• two propositions are logically equivalent if the Final Collums of their truth tables are identical.

↳ the propositions $p \wedge (q \vee r)$ and $(p \wedge q) \vee (p \wedge r)$ are logically equivalent

Tautology

- a tautology is a proposition that is true whatever the values of its variables. Similarly, a contradiction is a proposition that is always false

Tautology → True in every world → tells us nothing

contradiction → true in No worlds → tells us only falsehood

writing = For 'is equivalent to', equivalences with only a single variable include

proof: Simple case analysis. Construct truth tables with rows for $p = F$ and $p = T$

$p \vee p \equiv p$, $p \wedge p \equiv p$ } idempotence

$p \vee \neg p \equiv T$, $p \wedge \neg p \equiv F$ } excluded middle

$\neg\neg p \equiv p$ > double negation

$p \vee F \equiv p$, $p \wedge T \equiv p$ } identity

$p \vee T \equiv T$, $p \wedge F \equiv F$ } Strictness

proof: Case analysis. for each equivalence construct a truth table

• a convention of priority among connectives, together with the associativity of $\vee$ and $\wedge$, reduces the need for brackets.

$(p \vee q) \vee r \equiv p \vee (q \vee r)$, $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$ } associativity

$p \vee q \equiv q \vee p$, $p \wedge q \equiv q \wedge p$ } Commutativity

$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$, $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$ } distributivity

$\neg(p \vee q) \equiv (\neg p) \wedge (\neg q)$, $\neg(p \wedge q) \equiv (\neg p) \vee (\neg q)$ } de morgan

High priorities

$\neg$	not
$\wedge$	and
$\vee$	or
$\Rightarrow$	implies
$\Leftrightarrow$	if and only if

XOR, NOR, NAND

- There are 2^{2^n} logically distinct propositions over n variables

↳ Proof:

- With n variables, a truth table has 2^n rows, and in defining column there are two possible entries (F,T) in each of these rows.

XOR connective

P	q	P XOR q
F	F	F
F	T	T
T	F	T
T	T	F

Theorem 4

The three connectives $\wedge$, $\vee$ and $\neg$ are enough to express logical equivalents of all others.

Proof.

Case analysis. Give a suitable proposition for every possible result column. For example:

$$p \text{ XOR } q \equiv (p \vee q) \wedge \neg(p \wedge q)$$

Corollary 5

Just two connectives, $\neg$ with either $\vee$ or $\wedge$, are enough.

Proof.

$$p \vee q \equiv \neg(\neg p \wedge \neg q)$$

therefore, $p \text{ XOR } q \equiv \neg(\neg p \wedge \neg q) \wedge \neg(p \wedge q)$

Theorem 5

Just one connective, either $\downarrow$ or $\uparrow$ is enough to express all others.

Proof.

$\uparrow$ alone is enough since

$$\neg p \equiv p \uparrow p$$
$$p \vee q \equiv \neg p \uparrow \neg q$$

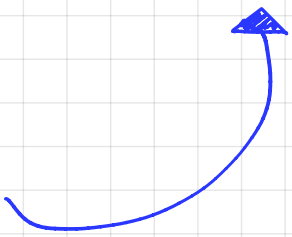
□

Ex. What about $\downarrow$?

Solution 6

We can use the mirror law to prove that $\downarrow$ alone is also enough

$$\neg p \equiv p \uparrow p$$
$$p \vee q \equiv \neg p \uparrow \neg q$$
$$p \wedge q \equiv \neg p \downarrow \neg q$$
$$\neg p \equiv p \downarrow p$$



Predicates + Quantifiers

- The implies connective can be defined by the truth table:

P	Q	P ⇒ Q
F	F	T
F	T	T
T	F	F
T	T	T

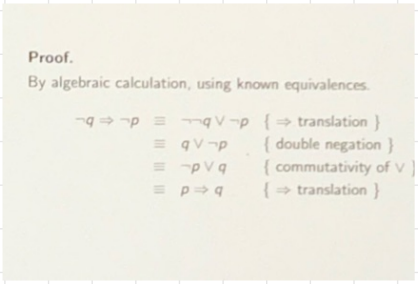
- or by the equivalence

$$P \Rightarrow Q \equiv \neg P \vee Q$$

- Intuitively $P \Rightarrow Q$ can be read 'if P then Q'

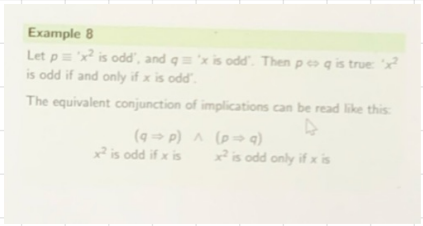
- for any implication $P \Rightarrow Q$

Converse $Q \Rightarrow P$;
inverse $\neg P \Rightarrow \neg Q$;
contrapositive $\neg Q \Rightarrow \neg P$;
original $P \Rightarrow Q \equiv \neg Q \Rightarrow \neg P$ contrapositive
converse $Q \Rightarrow P \equiv \neg P \Rightarrow \neg Q$ inverse



Proof.
By algebraic calculation, using known equivalences.

$\neg q \Rightarrow \neg p \equiv \neg\neg q \vee \neg p$ { ⇒ translation }
$\equiv q \vee \neg p$ { double negation }
$\equiv \neg p \vee q$ { commutativity of ∨ }
$\equiv p \Rightarrow q$ { ⇒ translation }

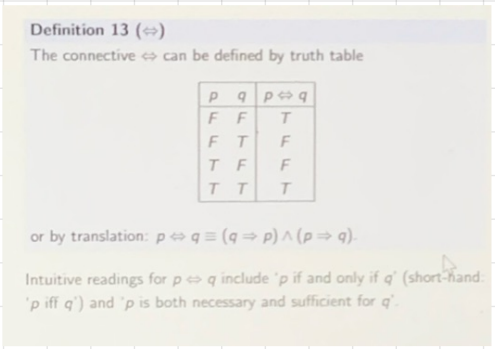


Example 8

Let $p \equiv$ 'x^2 is odd', and $q \equiv$ 'x is odd'. Then $p \Leftrightarrow q$ is true: 'x^2 is odd if and only if x is odd'.

The equivalent conjunction of implications can be read like this:

$(q \Rightarrow p) \wedge (p \Rightarrow q)$
x^2 is odd if x is $\quad$ x^2 is odd only if x is



Definition 13 (⇔)

The connective ⇔ can be defined by truth table

p	q	p ⇔ q
F	F	T
F	T	F
T	F	F
T	T	T

or by translation: $p \Leftrightarrow q \equiv (q \Rightarrow p) \wedge (p \Rightarrow q)$.

Intuitive readings for $p \Leftrightarrow q$ include 'p if and only if q' (short-hand: 'p iff q') and 'p is both necessary and sufficient for q'.

- a predicate is a proposition $p(v_1, v_2 \dots v_n)$ depending on variables v_1, v_2, v_n.
- Given a value for each V_i, p defines a statement that is true or false
  - The $\exists$ is the existential qualifier. A formula $\exists x, p(x)$ reads 'there exists a value of x such that $p(x) = T$'.
  - The symbol $\forall$ is the universal quantifier. A formula $\forall x, p(x)$ reads 'for all values of x, $p(x) = T$'
  - $\exists$ formula stands for a chain of statements linked by $\vee$
  - $\forall$ stands for a chain linked by $\wedge$

de Morgans law extends to quantifiers

$\neg(\exists x\, p(x)) \equiv \forall x, \neg p(x)$

$\neg(\forall x, p(x)) \equiv \exists x, \neg p(x)$

Proof – informal sketch only.

$\neg(\exists x, p(x)) \equiv$ 'no x makes $p(x)$ true'
$\equiv$ 'for every x, $p(x)$ is false'
$\equiv \forall x, \neg p(x)$

$\neg(\forall x, p(x)) \equiv$ 'not every x makes $p(x)$ true'
$\equiv$ 'for some x, $p(x)$ is false'
$\equiv \exists x, \neg p(x)$

Restricted quantifiers

- often the range of values for which a quantified statement holds is restricted:

$\forall x, (r(x) \Rightarrow p(x))$ 'for all x of type r, $p(x)$'

$\exists x, (r(x) \wedge q(x))$ 'for some of x of type r, $q(x)$'

↳ example

Q What does $\exists x, (student(x) \Rightarrow asleep(x))$ mean?

We can use the laws we know to unpack the statement:
$\neg student(x) \vee asleep(x)$
Not a student x or asleep x

$\neg(student(x) \wedge \neg asleep(x))$
Is not the case that x is a student and not asleep.

The best we can do would probably be "There is someone who is not an awake student."

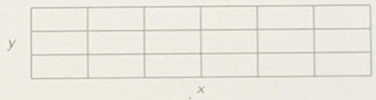
Theorem 8
Extended de Morgan also works with restricting predicates:

$\neg(\exists x, (r(x) \wedge p(x))) \equiv \forall x, (r(x) \Rightarrow \neg p(x))$
$\neg(\forall x, (r(x) \Rightarrow p(x))) \equiv \exists x, (r(x) \wedge \neg p(x))$

Proof of the first equivalence.

$\neg(\exists x, (r(x) \wedge p(x))) \equiv \forall x, \neg(r(x) \wedge p(x))$ { Theorem 7 }
$\equiv \forall x, (\neg r(x) \vee \neg p(x))$ { de Morgan's law }
$\equiv \forall x, (r(x) \Rightarrow \neg p(x))$ { $\Rightarrow$ translation }
□

In a grid of x, y values, suppose each square where $p(x, y)$ is true is shaded.



In (a), LHS ≡ RHS ≡ 'some square is shaded'.
In (b), LHS ≡ RHS ≡ 'every square is shaded'.
But in (c), LHS ≡ 'there is a column with every square shaded', whereas RHS ≡ 'in every row some square is shaded'.

When one quantifier follows another only like quantifiers commute:

(a) $\exists x, \exists y, p(x, y) \equiv \exists y, \exists x, p(x, y)$
(b) $\forall x, \forall y, p(x, y) \equiv \forall y, \forall x, p(x, y)$

! Be careful, when quantifiers are not the same, they do not commute:

(c) $\exists x, \forall y, p(x, y) \not\equiv \forall y, \exists x, p(x, y))$

Witness + Counter example

To prove a ∃ formula, or disprove a ∀ formula, a single value is enough.

For an existential formula ∃x,p(x) a witness is a value of x making p(x) true, so proving the truth of ∃x,p(x) as a whole.

counter-example

Definition 18 (counter-example)

For a universal formula $\forall x, p(x)$ a *counter-example* is a value of x for which $p(x)$ is false, exposing the falsehood of $\forall x, p(x)$ as a whole.

Q. **Fool's prime**: $\forall n, prime(3^n + 2)$ is a false claim. What is the smallest counter-example?

Solution 19

Indeed for $n = 0, 1, 2, 3, 4$ *the formula* $3^n + 2$ *yields primes* $3, 5, 11, 29, 83$. *The smallest counter-example is* $n = 5$ *for which the formula yields* $245 = 5 \times 7 \times 7$.

3/4

Natural induction

- To prove a ∀ formula, or disprove a ∃ formula, inductive arguments are often used.

- For a predicate p(n) about the natural numbers (n = 0, 1, 2, . . .) natural induction is a proof by the following argument.

p(0) base

p(k) ⇒ p(k+1) inductive case

∴ ∀n, p(n)

- When dealing with the inductive case, we assume p(k) and show that p(k + 1) follows.
- The technical term for the assumption p(k) is the inductive hypothesis.
- Sometimes p(0) has no useful meaning, and p(1) is the appropriate base case.

Nat. induction Continued

- When dealing with the inductive case, we assume p(k) and show that p(k + 1) follows.
- The technical term for the assumption p(k) is the inductive hypothesis.
- Sometimes p(0) has no useful meaning, and p(1) is the appropriate base case.

Proof by induction on n.

Base Case: when $n = 0$, it is true that $0 = 0(0+1)/2$

Inductive Case: when $n = k$, assume

$$0 + 1 + \ldots + k = k(k+1)/2$$

then

$$\begin{aligned} & 0 + 1 + \ldots + k + (k+1) \\ = \; & (0 + 1 + \ldots + k) + (k+1) \\ = \; & (k(k+1)/2) + (k+1) \text{ by assumption} \\ = \; & (k(k+1) + 2(k+1))/2 \\ = \; & (k+1)((k+1)+1)/2 \end{aligned}$$

Theorems about structures such as formulae, programs and circuits may also be proved by natural induction, where the inductive variable is some measure of size.

$\neg(P_1 \vee \ldots \vee P_n) \equiv \neg P_1 \wedge \ldots \wedge \neg P_n$

- Induction is often a good way to prove a ∀ claim or to disprove a ∃ one.
- Though basic induction is about numbers, induction can be used in proofs about any kinds of values with numeric sizes.

Sets

- a set is a collection of distinct members
- A set specification is an expression of the form {member | predicate}. Or elements can simply be listed, in any order, between {. . .}.
  - ↳ The set {p | prime(p)∧p<10} has elements 2,3,5,7. The same set can also be expressed as {7, 2, 5, 3}.

∈, set membership

$x \in S$ means x is a member of set S

$x \notin S$ means $\neg(x \in S)$

↳ Example :
Both $3 \in \mathbb{N}$ and $-1/\in \mathbb{N}$ are true, but $\pi \in \mathbb{Z}$ is false

Definition 22 (naturals N, integers Z)

N is the set of natural (whole non-negative) numbers {0,1,2,3. . . }; **Z** is the set of all whole numbers {. . . -2,-1,0,1,2. . . }.

Definition 23 (empty set Ø; universal set U)

The *empty set* Ø contains no element; that is, $\emptyset = \{x|F\}$. The *universal set* **U** contains all possible elements (in some agreed world of one or more values); that is, $\mathbf{U} = \{x|T\}$.

Definition 25 (axioms of membership)

$\{x|x \in S\} = S$ and $y \in \{x|p(x)\} \equiv p(y)$

Definition 26 (⊆, subset; ⊂, proper subset)

$P \subseteq Q$ 'P is a *subset* of Q' is true iff every member of P is also a member of Q. The stronger assertion $P \subset Q$ 'P is a *proper subset* of Q' holds if $P \subseteq Q$ but $P \neq Q$.

Example 18

For any set P: $\emptyset \subseteq P$, $P \subseteq P$ and $P \subseteq \mathbf{U}$ are all true. In the world of whole numbers $\{n|n > 2 \wedge prime(n)\} \subset \{n|odd(n)\}$.

The predicates $\subseteq$ and $=$ between sets can also be expressed in terms of logical connectives:
$P \subseteq Q \equiv \forall x, x \in P \Rightarrow x \in Q$ $P = Q \equiv \forall x, x \in P \Leftrightarrow x \in Q$

Q. How can $\subset$ be defined in logical terms?

$$\begin{aligned} P \subset Q &\equiv (P \subseteq Q) \wedge \neg(P = Q) \\ &\equiv (\forall x, x \in P \Rightarrow x \in Q) \wedge \neg(\forall x, x \in Q \Rightarrow x \in P) \\ &\equiv (\forall x, x \in P \Rightarrow x \in Q) \wedge (\exists x, x \in Q \wedge x \notin P) \end{aligned}$$

Intersection $\cap$: union $\cup$

If P, Q are sets, their
union $P \cup Q = \{x | x \in P \vee x \in Q\}$ intersection $P \cap Q = \{x | x \in P \wedge x \in Q\}$

Difference $\setminus$

If P, Q are sets, their
difference $P \setminus Q = \{x | x \in P \wedge x \notin Q\}$

Idempotence, strictness, identity

Theorem 11

For any set $P \subseteq$ universe $\mathbf{U}$:

$$\left.\begin{aligned} P \cap P &= P \\ P \cup P &= P \end{aligned}\right\} \textit{idempotence}$$

$$\left.\begin{aligned} P \cap \emptyset &= \emptyset \\ P \cup \mathbf{U} &= \mathbf{U} \end{aligned}\right\} \textit{strictness}$$

$$\left.\begin{aligned} P \cap \mathbf{U} &= P \\ P \cup \emptyset &= P \end{aligned}\right\} \textit{identity}$$

Proof.
Since $x \in \emptyset \equiv F$ and $x \in \mathbf{U} \equiv T$ for any element x, these laws follow from the corresponding laws about $\wedge$ and $\vee$. For example:

$$\begin{aligned} & P \cap \mathbf{U} \\ = \ & \text{by definition of } \cap \\ & \{x | x \in P \wedge x \in \mathbf{U}\} \\ = \ & \text{by definition of } \mathbf{U} \\ & \{x | x \in P \wedge T\} \\ = \ & \text{by identity law for } \wedge \\ & \{x | x \in P\} \\ = \ & \text{by the axioms of } \in \\ & P \end{aligned}$$

Theorem 12

For any sets P,Q,R

$$\left.\begin{aligned} P \cap Q &= Q \cap P \\ P \cup Q &= Q \cup P \end{aligned}\right\} \textit{commutativity}$$

$$\left.\begin{aligned} P \cap (Q \cap R) &= (P \cap Q) \cap R \\ P \cup (Q \cup R) &= (P \cup Q) \cup R \end{aligned}\right\} \textit{associativity}$$

$$\left.\begin{aligned} P \cap (Q \cup R) &= (P \cap Q) \cup (P \cap R) \\ P \cup (Q \cap R) &= (P \cup Q) \cap (P \cup R) \end{aligned}\right\} \textit{distributivity}$$

Proof.
By appeal to the corresponding laws about $\wedge$ and $\vee$. For example:

$$\begin{aligned} & P \cap Q \\ = \ & \text{by definition of } \cap \\ & \{x | x \in P \wedge x \in Q\} \\ = \ & \text{by commutativity of } \wedge \\ & \{x | x \in Q \wedge x \in P\} \\ = \ & \text{by definition of } \cap \\ & Q \cap P \end{aligned}$$

□

Complement

Definition 29 (complement)

If P is a set, its

$$\text{COMPLEMENT} \quad P' = \{x | x \notin P\}$$

complement

Theorem 13

Within a universe $\mathbf{U}$, $P' = \mathbf{U} \setminus P$.

Proof.

$$\begin{array}{rcll} P' & = & \{x | x \notin P\} & \text{by definition} \\ & = & \{x | T \wedge x \notin P\} & \text{by identity law for } \wedge \\ & = & \{x | x \in \mathbf{U} \wedge x \notin P\} & \text{by definition of } \mathbf{U} \\ & = & \mathbf{U} \setminus P & \text{by definition of } \setminus \end{array}$$

□

7/12

complement

Theorem 14

$$\left.\begin{array}{rcl} P \cap P' & = & \emptyset \\ P \cup P' & = & \mathbf{U} \end{array}\right\} \textit{excluded middle}$$

$$\left.\begin{array}{rcl} (P \cap Q)' & = & P' \cup Q' \\ (P \cup Q)' & = & P' \cap Q' \end{array}\right\} \textit{de Morgan}$$

Proof.

Same technique as Theorems 11–13. □

8/12

Set equality

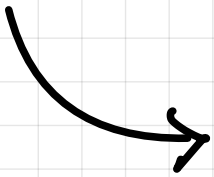
Besides calculating set equalities using such laws, here's a commonly used argument that splits an equality proof into two parts:

$$\frac{R \subseteq S \qquad S \subseteq R}{\therefore R = S}$$

A common way to show R ⊆ S is to reason that, for any choice of x ∈ R we can guarantee x ∈ S. Similarly for S ⊆ R.

Example 20 (Theorem 14 revisited)

The law $P \cup P' = \mathbf{U}$ could be shown as follows.
$\subseteq$: Certainly, since every set $\subseteq \mathbf{U}$.
$\supseteq$: Choose any $x \in \mathbf{U}$. Either $x \in P$ or $x \notin P$. If $x \in P$ then $x \in (P \cup \ldots)$, so $x \in (P \cup P')$. If $x \notin P$ then $x \in P'$, so $x \in (\ldots \cup P')$, and again $x \in (P \cup P')$. □



Suppose one of the sets is empty:

$$\frac{R \subseteq \emptyset \qquad \emptyset \subseteq R}{\therefore R = \emptyset}$$

Since $\emptyset \subseteq R$ is always true, we only need to show $R \subseteq \emptyset$: if $x \in R$ then $x \in \emptyset$. That is, we show that $x \in R$ is always false, whatever x may be.

Example 21 (Theorem 14 revisited again)

The law $P \cap P' = \emptyset$ fits this pattern. Suppose $x \in (P \cap P')$. Then $x \in P$. But also $x \in P'$; that is, $x \notin P$. By the excluded-middle law we know $x \in P \wedge x \notin P$ is false. So there can be no such x. That is, $P \cap P' = \emptyset$. □

Set Cardinality

Definition 30 (#, set cardinality)

If S is a set, $\#S$ (or $card(S)$, or $|S|$) expresses its *cardinality*, the number of elements it has.

Example 22

$\#\emptyset = 0$, $\#\{0,1,2\} = 3$.

- Unlike the operators producing sets, # has no immediate counterpart in predicate logic, but here is a useful rule.
- For any set S and predicate p
  #S = #{x|x ∈ S ∧ p(x)} + #{x|x ∈ S ∧ ¬p(x)}

  Proof by the laws of excluded middle.
  Since for any x ∈ S, p(x) ∨ ¬p(x) is true the sum does not omit any element; and since p(x) ∧ ¬p(x) is false no element is counted twice.

Definition 31 (disjoint sets)

Sets P, Q are **disjoint** if they have no elements in common.

Example 23

The sets of odd and even numbers are disjoint. More generally, S and S' are disjoint for any set S.

Theorem 16

The following are all equivalent ways of asserting that P and Q are disjoint:

(a) $P \subseteq Q'$

(b) $P \setminus Q = P$

(c) $P \cap Q = \emptyset$

(d) $P \cup Q' = Q'$

Theorem 17

If P and Q are finite sets:

(a) $\quad Q \subseteq P' \Rightarrow \#(P \cup Q) = \#P + \#Q$

Proof.

If $Q \subseteq P'$ then

$\#(P \cup Q)$

$=$ by Theorem 15

$\#\{x|x\in(P\cup Q)\wedge x\in P\} + \#\{x|x\in(P\cup Q)\wedge x\notin P\}$

$=$ $\#\{x|x \in P\} + \#\{x|x \in Q\}$ since $Q \subseteq P'$

$=$ $\#P + \#Q$ by axiom of membership

□

Theorem 17 (continued.)

If P and Q are finite sets:

(b) $\quad Q \subseteq P \Rightarrow \#(P \setminus Q) = \#P - \#Q$

Proof.

If $Q \subseteq P$ then

$\#(P \setminus Q) + \#Q$

$=$ by definition of $\setminus$ and axiom of membership

$\#\{x|x \in P \wedge x \notin Q\} + \#\{x|x \in Q\}$

$=$ since $Q \subseteq P$

$\#\{x|x \in P \wedge x \notin Q\} + \#\{x|x \in P \wedge x \in Q\}$

$=$ $\#P$ (by Theorem 15)

□

Disjoint unions

Definition 32 (⊎, disjoint union)
$P \uplus Q$ is the union of disjoint sets P and Q.

So from Theorem 17, $\#(P \uplus Q) = \#P + \#Q$.

Theorem 18 (union decomposition)

$$P \cup Q = (P \setminus Q) \uplus (P \cap Q) \uplus (Q \setminus P)$$

Proof.
Suppose $x \in P \cup Q$; let $p \equiv x \in P, q \equiv x \in Q$. Then $p \vee q$ is true.
From the $\vee$ truth table:
(1) $p = T, q = F$, so $x \in P \setminus Q$, or
(2) $p = T, q = T$, so $x \in P \cap Q$, or
(3) $p = F, q = T$, so $x \in Q \setminus P$.

Theorem 19 (cardinality rule)

$$\#(P \cup Q) + \#(P \cap Q) = \#P + \#Q$$

Proof.

$$\begin{aligned} & \#(P \cup Q) + \#(P \cap Q) \\ = \; & \#((P \setminus Q) \uplus (P \cap Q) \uplus (Q \setminus P)) + \#(P \cap Q) \\ = \; & \#(P \cap Q) + \#(P \setminus Q) + \#(Q \cap P) + \#(Q \setminus P) \\ = \; & \#P + \#Q \end{aligned}$$

Powersets

Definition 33 (powerset)
If S is a set, its *powerset* 2^S is the set whose elements are all the subsets of S. That is:

$$2^S = \{R \mid R \subseteq S\}$$

Q. Why the name powerset and notation 2^S?

Answer
$\#2^S = 2^{\#S}$

Example 24
If S is the set $\{ham, cheese, tomato\}$ then 2^S is the set:

$$\left\{ \begin{array}{llll} \{ & ham & , cheese & , tomato \} , \\ \{ & ham & , cheese & \} , \\ \{ & ham & , & tomato \} , \\ \{ & ham & & \} , \\ \{ & & cheese & , tomato \} , \\ \{ & & cheese & \} , \\ \{ & & & tomato \} , \\ & \emptyset & & \end{array} \right\}$$

Partition

Definition 34 (partition)
A set $\{P_1, P_2, P_3, \ldots\}$ of non-empty subsets of a set S is a *partition* of S exactly if:
(1) $P_1 \cup P_2 \cup P_3 \cup \ldots = S$ (P_i's cover S), and
(2) $j \neq k \Rightarrow P_j \cap P_k = \emptyset$ (P_i's are disjoint).

The only partition of $\emptyset$ is itself $\emptyset$. If P is a partition of $S \neq \emptyset$ then $1 \leq \#P \leq \#S$.

Example 25
Let $S = \{A, B, C\}$. There are five possible partitions of S.

A single-subset partition, all elements together:

$$\{\{A, B, C\}\}$$

Two in one subset, and one in another:

$$\begin{array}{rl} & \{\{A, B\}, \{C\}\} \\ \text{or} & \{\{A, C\}, \{B\}\} \\ \text{or} & \{\{B, C\}, \{A\}\} \end{array}$$

Each element alone in a subset:

$$\{\{A\}, \{B\}, \{C\}\}$$

Theorem 20
If $\{P_1, P_2, P_3, \ldots\}$ is a partition of S, and $e \in S$, then $\exists! i, e \in P_i$.

Proof.
The covering property guarantees $\exists i$, and disjointness guarantees that i is unique.

Theorem 21
Let $parts(n)$ be the number of distinct partitions of a set with n elements.

$$\begin{aligned} parts(0) &= 1 \\ parts(n) &= \sum_{j=0}^{n-1} (C(n-1, j) \times parts(n-j-1)) \end{aligned}$$

where $C(n, m) = \frac{n!}{m!(n-m)!}$ is the number of ways of choosing an m-element subset from an n element set.

Topic 1 Lecture 4 again!

Set products

- The cartesian product A × B of two sets A, B is the set of ordered pairs {(a, b)|a ∈ A ∧ b ∈ B}.

Example 26

If $D = \{Mon, Tue, Wed, Thu, Fri\}$ and $T = \{\text{AM}, \text{PM}\}$, then $D \times T =$ is the set

$$\{(Mon, \text{AM}), \quad (Tue, \text{AM}), \quad \cdots \quad (Fri, \text{AM}),$$
$$(Mon, \text{PM}), \quad (Tue, \text{PM}), \quad \cdots \quad (Fri, \text{PM})\}$$

of all ten (weekday, time) pairs.

Binary Relation

- A binary relation between two sets A, B is a subset of A × B. If R is such a relation, we write R : A ↔ B. Also, aRb is short for (a,b) ∈ R.

Source + Target

If R : A ↔ B then A is the source of R and B is the target of R. When the source and target set are the same, say R : A ↔ A, we call R a relation on A.

Example 28

For sets D, T as in Example 26, the following $D \leftrightarrow T$ relation represents the timetable for some regular event.

$$\{(Mon, \text{PM}), (Tue, \text{PM}), (Wed, \text{AM})\}$$

Example 29

If $A = \{1, 2, 3, 4\}$ the $<$ relation on A is $\{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$.

For instance, $1 < 3$ is true, but $4 < 2$ is false.

Domain + Range

- The domain of R : A ↔ B is a subset of A. It contains every a ∈ A for which there is at least one (a,...) pair in R.
- The range of R : A ↔ B is a subset of B. It contains every b ∈ B for which there is at least one (...,b) pair in R.

Example 30

For the timetable relation

$$\{(Mon, \text{PM}), (Tue, \text{PM}), (Wed, \text{AM})\}.$$

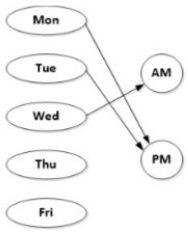
the domain is $\{Mon, Tue, Wed\}$ and the range is $\{\text{AM}, \text{PM}\}$.

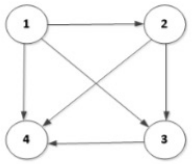
And for the $<$ relation

$$\{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$$

the domain is $\{1, 2, 3\}$ and the range is $\{2, 3, 4\}$.

Digraphs





- A directed graph is a collection of nodes (or vertices), some pairs of which are connected by arcs (or arrows).

↳ To illustrate a relation R : A ↔ B as a digraph:
If A and B are not the same, draw nodes first for each element in A then for each element in B. If A and B are the same, draw just a single node for each element.
Draw an arc from the a node to the b node wherever aRb.

Adjacency matrix

- Let R : A ↔ B. The adjacency matrix MR has #A rows and #B columns. Writing mjk for the element in the jth row and kth column of MR: mjk = 1 if ajRbk, and mjk = 0 otherwise.

$$\begin{array}{c} \\ \textit{Mon} \\ \textit{Tue} \\ \textit{Wed} \\ \textit{Thu} \\ \textit{Fri} \end{array} \begin{array}{c} \begin{array}{cc} \textit{AM} & \textit{PM} \end{array} \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

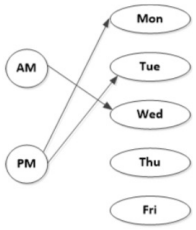
Note: It matters how elements of A and B are assigned to columns and rows, but they are not usually labelled.

5/10

Relational inverse

- If R : A ↔ B then its inverse R–1 : B ↔ A is defined by the rule bR–1a ⇔ aRb.

The inverse of the relation $\{(Mon, \text{PM}), (Tue, \text{PM}), (Wed, \text{AM})\}$ is the relation $\{(\text{PM}, Mon), (\text{PM}, Tue), (\text{AM}, Wed)\}$.



matrix transpose

- If M is a v ×u matrix its transpose MT is a u×v matrix whose columns are the rows of M.
- So the adjacency matrix for R–1 is the transpose of the adjacency matrix for R.

! Don't confuse inverse with complement.
When taking the complement of a relation $R : A \leftrightarrow B$ the appropriate universe **U** is $A \times B$.
The complement $R' : A \leftrightarrow B$ is specified by

$$aR'b \;\Leftrightarrow\; \neg(aRb).$$

! **Example 29 revisited**
Consider again the relation $<$:

$<^{-1} = >$ because $b > a \Leftrightarrow a < b$
$<' = \geq$ because $a \geq b \Leftrightarrow \neg(a < b)$

- Binary relations can also be represented as directed graphs, or as adjacency matrices.
- Inverting a relation inverts the order of pairs in the set, or the direction of arcs in the digraph, or the roles of rows and columns in the matrix.

Function

- R :A↔B is called a function if for every a ∈ A there is at most oneb∈B suchthat(a,b)∈R.
  ∀a ∈ A,#{b | (a,b) ∈ R} ≤ 1
- matrix: at most one 1 in each row
- digraph: at most one arc from each A-node

The relation {(Mon, pm), (Tue, pm), (Wed, a
a function.
The < relation on {1, 2, 3, 4} is not a functio
instance, 1 < 2 but also 1 < 3 and 1 < 4.

Total

R : A ↔ B is total if for every a ∈ A there is at least one b∈B suchthat(a,b)∈R.

∀a ∈ A,#{b | (a,b) ∈ R} ≥ 1
domain = source
matrix: at least one 1 in each row digraph: at least one arc from each A-node

The relation $\{(Mon, \text{PM}), (Tue, \text{PM}), (Wed, \text{AM})\}$ with source D is not total, as *Thu* and *Fri* are not in the domain.

Neither is $<$ on A total, as 4 is not in the domain — there is no element $n \in \{1, 2, 3, 4\}$ such that $4 < n$.

When $f : A \leftrightarrow B$ is a total function, we write $f : A \to B$. For each $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$, and we may express b as $f(a)$.

Injective

R : A ↔ B is injective if for every b ∈ B there is at most one a ∈ A such that (a,b) ∈ R.

$$\forall b \in B, \#\{a \mid (a, b) \in R\} \leq 1$$

MATRIX: at most one 1 in each column

DIGRAPH: at most one arc to each B-node

Bijection

A total function that is both injective and surjective is called a 1-1 correspondence, or a bijection.

Surjective

R : A ↔ B is surjective if for every b ∈ B there is at least one a ∈ A such that (a,b) ∈ R.

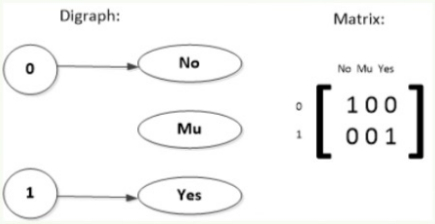
∀b ∈ B,#{a | (a,b) ∈ R} ≥ 1
range = target
matrix: at least one 1 in each column
digraph: at least one arc to each B-node

	Out Degree	In Degree	Comments
Function	$\leqslant 1$	-	
Total	$\geqslant 1$	-	Domain = Source
Total function	$= 1$	-	$f : A \to B$
Injective	-	$\leqslant 1$	
Surjective	-	$\geqslant 1$	Range = Target
Bijective	$=1$	$=1$	$1-1$ Correspondence

injective, surjective, and bijective

Example 31

Given the sets $Bits = \{0, 1\}$ and $Reps = \{No, Mu, Yes\}$, let $f : Bits \to Reps$ be the total function $\{(0, No), (1, Yes)\}$. It is injective, but not surjective.

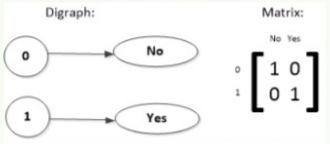


2/10

injective, surjective, and bijective

Example 31 (continued.)

But $f : Bits \to (Reps \setminus \{Mu\})$ is a bijection.



Remember the definition of the inverse relation:
If $R : A \leftrightarrow B$ then its *inverse* $R^{-1} : B \leftrightarrow A$ is defined by the rule $bR^{-1}a \Leftrightarrow aRb$.

Then note that the inverse of a function f is not always a function.

Theorem 22

For any function f:

f^{-1} *is a function* $\Leftrightarrow$ f *is injective*

f^{-1} *is total* $\Leftrightarrow$ f *is surjective*

Permutation

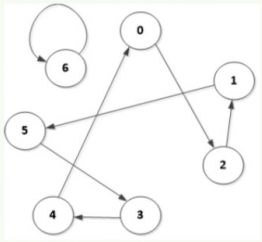
- A permutation is a 1-1 correspondence on a finite set.

permutation

Example 32

Let $X = \{0, 1, 2, 3, 4, 5, 6\}$ and define $f : X \to X$ by $f(x) = (3x + 2) \bmod 7$.

Digraph:



5/10

permutation

Example 32 (continued.)

Matrix:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

6/10

Q. In terms of $\#S$, how many different permutations of S are there?

There are $(\#S)!$ permutations $S \to S$

Corollary 33

for r-permutations of elements of a set of cardinality $\#S$ that preclude repetition: $P(\#S, r) = \frac{\#S!}{(\#S-r)!} = (\#S)!$
where $r = (\#S)$

Proof of f^{-1} is a function $\Leftrightarrow$ f is injective.

Let $f : A \leftrightarrow B$. Then:

f is injective $\Leftrightarrow$ by definition of injective
$\forall b \in B, \#\{a \mid (a, b) \in f\} \leq 1$
$\Leftrightarrow$ by definition of inverse
$\forall b \in B, \#\{a \mid (b, a) \in f^{-1}\} \leq 1$
$\Leftrightarrow$ by definition of a function
f^{-1} is a function

Relational Composition

- For any R : A ↔ B and S : B ↔ C the relational composition S ∘ R : A ↔ C is defined by :
  (a, c) ∈ S ∘ R ⇔ ∃b, (a, b) ∈ R ∧ (b, c) ∈ S

Theorem 23

- If f : A ↔ B and g : B ↔ C are both functions, so is g ∘ f :A↔C.

Proof.
Suppose $(a, c_1) \in g \circ f$ and $(a, c_2) \in g \circ f$. Let b_1 and b_2 be witnesses. Q. To what?
So $(a, b_1) \in f$ and $(a, b_2) \in f$, and as f is a function $b_1 = b_2$.
But also $(b_1, c_1) \in g$ and $(b_2, c_2) \in g$. So, as $b_1 = b_2$ and g is a function, $c_1 = c_2$ as required. □

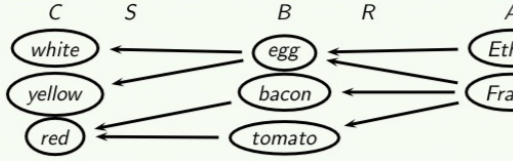
Theorem 24

- If f : A↔B and g : B ↔C are both (i) total, or (ii) surjective, or (iii) injective, then so is their composition g ∘ f : A ↔ C .

Proof.
(i) Suppose $a \in A$. Then as f is total $\exists b \in B, (a, b) \in f$ and as g is total $\exists c \in C, (b, c) \in g$. So $\exists c \in C, (a, c) \in g \circ f$ as required. □

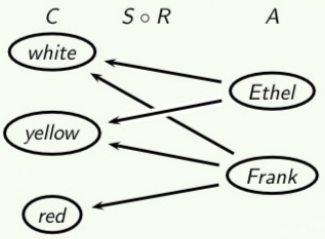
Corollary 25
If f ,g are both 1-1 correspondences then so is g ∘ f . In other words, If f ,g are both bijections then so is g ∘ f .

Example 33 (Ethel and Frank's breakfast)



elational composition

Example 33 (Ethel and Frank's breakfast continued



Intuition: $aRb \Leftrightarrow$ 'a eats b', $bSc \Leftrightarrow$ 'b is c in colour' a
$a(S \circ R)c \Leftrightarrow$ 'a eats something c in colour'.

Boolean sum + product

- Definition 50 (Boolean sum and product)
  LetB bethesetofBooleanvalues{0,1}. Ifb1 ∈B andb2 ∈B then their:
  - sumb1+b2 = max(b1,b2)
  - productb1·b2 = min(b1,b2)
  - So + and · are just like ∨ and ∧ but take 0 and 1 as operands, instead of F and T.

We can use these operations to define a product for matrices of Boolean values.

Definition 51 (matrix product)
If $F : w \times v$ and $G : v \times u$ are matrices of Boolean values, their *matrix product* $H = F \bullet G$ is a $w \times u$ Boolean matrix with elements defined by

$$h_{ki} = \sum_{j=1}^{v} f_{kj} \cdot g_{ji}$$

where $\sum$ means a chain of BOOLEAN +'s.

So the value in row k, column i of $F \bullet G$ is derived from row k of F and column i of G.

Relation Composition + matrix product

If R,S are composable relations with matrices MR,MS then MS ∘ R =MR •MS.

relation composition and matrix product

Proof.

If $R : A \leftrightarrow B$, $S : B \leftrightarrow C$ then $M_R \bullet M_S$ indeed has $\#A$ rows and $\#C$ columns, so it must be the adjacency matrix of SOME relation $\boxed{?} : A \leftrightarrow C$.

To show $\boxed{?} = S \circ R$:

$$
\begin{aligned}
& (a_i, c_k) \in \boxed{?} \\
\Leftrightarrow\ & (M_R \bullet M_S)_{ik} = 1 \\
\Leftrightarrow\ & \textstyle\sum_{j=1}^{v} (M_R)_{ij} \cdot (M_S)_{jk} = 1 \\
\Leftrightarrow\ & \exists j, ((M_R)_{ij} = 1) \wedge ((M_S)_{jk} = 1)) \\
\Leftrightarrow\ & \exists j, (a_i, b_j) \in R \wedge (b_j, c_k) \in S \\
\Leftrightarrow\ & (a_i, c_k) \in S \circ R
\end{aligned}
$$

Theorem 26

If the matrix product $F \bullet G$ is defined, then the product $G^T \bullet F^T$ is also defined (note reverse order) and $(F \bullet G)^T = G^T \bullet F^T$.

Corollary 27

The relational counterpart is the law $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.

Theorem 27

The matrix product operator $\bullet$ is associative: for any matrices $F: u \times v$, $G: v \times w$ and $H: w \times x$, $(F \bullet G) \bullet H = F \bullet (G \bullet H)$.

Corollary 28

The relational counterpart is the law $(T \circ S) \circ R = T \circ (S \circ R)$.

So we can write $F \bullet G \bullet H$ or $T \circ S \circ R$ without ambiguity.

Q. Are • (and ∘) *commutative* too?

No. Matrix Product (•) and Composition (∘) are NOT commutative.

Matrix Sum

- If F, G are both u×v matrices, their sum F +G is also a u×v matrix with elements (F + G)ij = Fij + Gij .
  So matrix + is commutative and associative.
- Theorem 28
  if R,S are both A ↔ B relations, the adjacency matrix MR∪S =MR +MS.

Example 35

Let R be the $<$ relation, and S the $=$ relation, on the set $A = \{1, 2, 3\}$. Then $R \cup S$ is the $\leq$ relation.

$$
\underset{M_<}{\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}} + \underset{M_=}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \underset{M_\leq}{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}}
$$

- As a matrix, a composition is a Boolean matrix product.
- Matrix product (•) and functional composition (∘) are associative.
- Matrix Product (•) and Composition (∘) are NOT commutative.
- matrix + is commutative and associative.

Summary

Relations

- The set product A × B contains all ordered pairs of an element from A with an element from B.
- A binary relation is a subset of the product of its source and target sets.
- Binary relations can also be represented as directed graphs, or as adjacency matrices.
- Inverting a relation inverts the order of pairs in the set, or the direction of arcs in the digraph, or the roles of rows and columns in the matrix.

Classification

- A relation that includes all source elements at most/least once is functional/ total.
- A relation that includes all target elements at most/least once is injective/ surjective.
- So a relation is functional/total exactly when its inverse is injective/surjective.

Composition

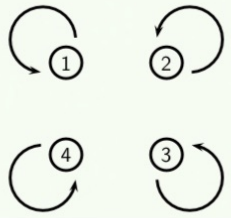
- Binary relations can composed only when the source of one is the target of the other.
- As a digraph, a composition has an arc where there is a two-arc path combining an arc from each relation in turn.
- As a matrix, a composition is a Boolean matrix product.

Identity Relations

- The identity relation IA on a set A is defined by a1IAa2 ⇔ (a1 =a2). If #A = n the adjacency matrix for IA is called In; it has 1's on the main diagonal and 0's everywhere else.

Example 36

The identity relation on $\{1,2,3,4\}$:



$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Powers of Relation

- Given R:A↔A, we define relations Rn for all n≥1: by R 1 = R , and ∀k ≥ 1, R k +1 = R ∘ R k .
- digraph: a1Rna2 exactly if in the digraph for R there is an n-arc path from a1 to a2.
- matrix: M 1 = M and ∀k ≥ 1, M k +1 = M • M k . (So by
- induction if R : A ↔ A has adjacency matrix MR then Rn has matrix (MR)n.)

Reflexive, irreflexive

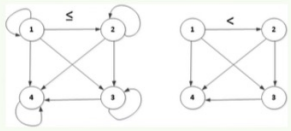
R : A ↔ A is reflexive if ∀a ∈ A,(a,a) ∈ R.
matrix: main diagonal is all 1's;
digraph: every node has a looping arc;

R is irreflexive if ∀a ∈ A,(a,a)/∈ R.
matrix: main diagonal is all 0's;
digraph: no node has a looping arc;

R is reflexive if R0 ⊆ R.

Example 37

On the set $\{1,2,3,4\}$ the $\leq$ relation is reflexive, but $<$ is irreflexive.



$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Symmetric / antisymmetric

Definition 56 (symmetric)

A relation $R : A \leftrightarrow A$ is *symmetric* if
$\forall a_1, a_2 \in A, (a_1, a_2) \in R \Rightarrow (a_2, a_1) \in R.$

MATRIX: symmetric about main diagonal;
DIGRAPH: all connections are two-way.

Definition 57 (antisymmetric)

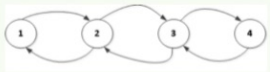
R is *antisymmetric* if
$\forall a_1, a_2 \in A, (a_1, a_2) \in R \wedge (a_2, a_1) \in R \Rightarrow (a_1 = a_2).$

MATRIX: any off-diagonal 1 is mirrored by 0;
DIGRAPH: any two-way connection is a loop.

Example 38

On the set $\{1,2,3,4\}$ the relation R with $mRn \Leftrightarrow |m-n|=1$ is symmetric:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



Example 37 revisited: $\leq$ is antisymmetric since if both $m \leq n$ and $n \leq m$ then $m = n$.
$<$ is also antisymmetric as $m < n \wedge n < m \equiv F$, and $F \Rightarrow \ldots$ is true.

Corollary 58

R is symmetric if $R^{-1} \subseteq R$.

Theorem 29

If $R : A \leftrightarrow A$ is both symmetric and antisymmetric then $R \subseteq I_A$.

Proof.

Let $r \in R$. Then $r = (a_1, a_2)$ for some $a_1 \in A$, $a_2 \in A$. Since R is symmetric $(a_2, a_1) \in R$ also. But then as R is antisymmetric $a_1 = a_2$. So $r \in I_A$. □

Transitive Relation

Theorem 30

A relation $R : A \leftrightarrow A$ is transitive IF AND ONLY IF $R^2 \subseteq R$.

A relation R : A ↔ A is transitive if
∀a1,a2,a3∈A,a1Ra2 ∧a2Ra3 ⇒a1Ra3.

digraph: wherever there is a two-arc path between nodes, there is also a direct arc between them in the same direction;
matrix: not obvious! (See Theorem 30.)

Proof.

IF part: Suppose $R^2 \subseteq R$. Then

$a_1 R a_2 \wedge a_2 R a_3$
$\Rightarrow \quad a_1 R^2 a_3$
$\Rightarrow \quad a_1 R a_3$

so R is transitive as required.

ONLY IF part: Suppose R is transitive. Then

$a_1 R^2 a_3$
$\Rightarrow \quad \exists a_2, a_1 R a_2 \wedge a_2 R a_3$
$\Rightarrow \quad a_1 R a_3$

so $r \in R^2 \Rightarrow r \in R$, that is $R^2 \subseteq R$ as required.

- A relation is transitive if it includes its own square

Example 37 revisited

That $<$ on $\{1, 2, 3, 4\}$ is transitive has already been noted from its digraph. Now $M_<$ and $(M_<)^2$ are:

$$M_< = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (M_<)^2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Indeed, these matrices show that $<^2 \subseteq <$.

5,

Example 38 revisited

By way of contrast, here are the corresponding matrices for the relation R defined by $mRn \Leftrightarrow |m - n| = 1$:

$$M_R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad (M_R)^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

So this R is NOT transitive: in fact, $R^2 \subseteq R'$.

6

Orderings

An ordering on a set A is a relation $\preccurlyeq$: A ↔ A that is reflexive, anti-symmetric and transitive.

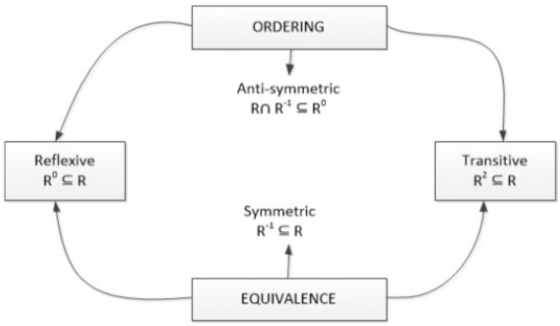
Example 39

Let $A = 2^{\mathbf{U}}$ for some universe $\mathbf{U}$. Then $\subseteq$ is an ordering on A.

REFLEXIVE: since $\forall s \in A, s \subseteq s$

ANTI-SYMMETRIC: $s_1 \subseteq s_2 \wedge s_2 \subseteq s_1 \Rightarrow s_1 = s_2$

TRANSITIVE: $s_1 \subseteq s_2 \wedge s_2 \subseteq s_3 \Rightarrow s_1 \subseteq s_3$



Strict ordering

A strict ordering is a relation that is irreflexive, anti-symmetric and transitive.

Example 40

Again let $A = 2^{\mathbf{U}}$. Then $\subset$ is a strict ordering on A:

IRREFLEXIVE: since $s \subset s \equiv F$

ANTI-SYMMETRIC: as $s_1 \subset s_2 \wedge s_2 \subset s_1 \equiv F$

TRANSITIVE: $s_1 \subset s_2 \wedge s_2 \subset s_3 \Rightarrow s_1 \subset s_3$

Theorem 31

If $\preceq$: $A \leftrightarrow A$ is an ordering, then $\preceq \setminus I_A$ is a strict ordering. And i
$\prec$: $A \leftrightarrow A$ is a strict ordering then $\prec \cup I_A$ is an ordering.

Hasse diagrams

Definition 60 (Hasse diagram)

A *Hasse diagram* shows an ordering. It is a digraph with some information left implicit.

REFLEXIVE: loops are omitted

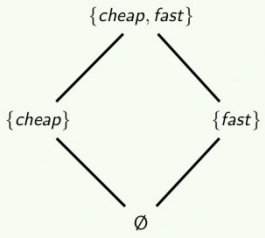
ANTI-SYMMETRIC: upward lines, not arrows

TRANSITIVE: arcs only if no longer path

Example 41

Consider two desirable properties of a computer, being cheap and being fast.

Diagram of $\subseteq$ over $2^{\{cheap, fast\}}$:



Definition 61 (comparable, incomparable)

For any ordering $\preceq$ on a set A, we say elements $a_1, a_2 \in A$ are *comparable* if $a_1 \preceq a_2 \vee a_2 \preceq a_1$. Otherwise a_1, a_2 are *incomparable*.

Example 41 revisited

Under the $\subseteq$ ordering the sets $\emptyset$ and $\{cheap, fast\}$ are comparable, but the sets $\{fast\}$ and $\{cheap\}$ are incomparable.

Definition 62 (total/linear ordering)

A *total* (or *linear*) ordering is one in which every pair of elements is comparable.

Definition 63 (partial ordering)

The term *partial ordering* describes an ordering that is not total. (It is also used to refer to orderings in general, to emphasize that they are not necessarily total.)

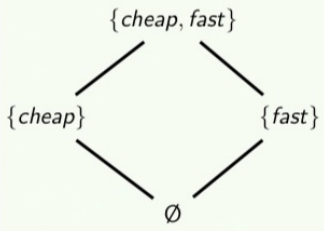
≤ is a total ordering on the set {1, 2, 3, 4} as any two numbers i, j are comparable: either i ≤ j or j ≤ i. But ⊆ is a partial ordering, not total, on the set 2{cheap,fast} as the subsets {cheap} and {fast} are incomparable: we have neither {cheap} ⊆ {fast} nor {fast} ⊆ {cheap}.

A partial ordering on a finite set A can always be extended to make a total ordering on A.

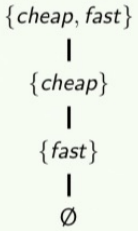
Example 42

Let $\preceq_1$ and $\preceq_2$ be the orderings with Hasse digrams:

$\preceq_1$:



$\preceq_2$:

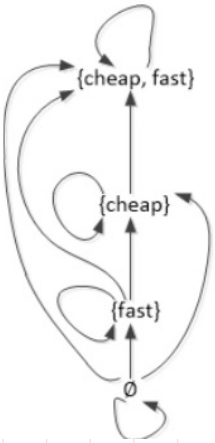


The $\preceq_1$ ordering is not total, but $\preceq_1 \subseteq \preceq_2$ which is a total ordering.

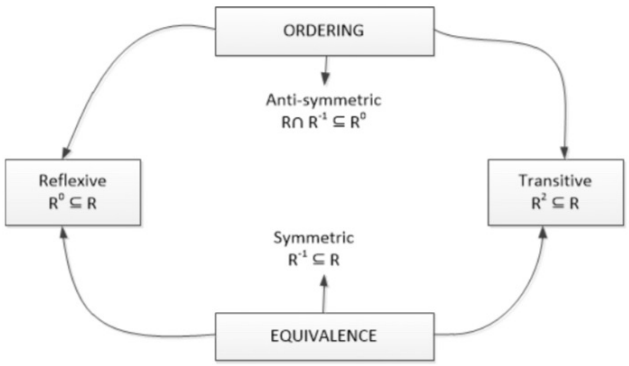
11/13

- A reflexive, antisymmetric and transitive relation is an ordering.
- Hasse diagrams show orderings.
- A strict ordering is a relation that is irreflexive, anti-symmetric and transitive.
- A total (or linear) ordering is one in which every pair of elements is comparable.

Q. If $\preceq_1 \subseteq \preceq_2$ why does the $\preceq_1$ diagram have more arcs than the $\preceq_2$ diagram?



1



equivalance class

If R≡ on A is an equivalence relation, for each a ∈ A we define the equivalence class [a]R≡ as the subset:
[a]R≡ = {a1 | aR≡a1}

Equivalance Relation

Definition 64 (equivalence relation)

A relation $R_{\equiv} : A \leftrightarrow A$ is an *equivalence* if it is reflexive, symmetric and transitive.

Example 43

Let $S_{\equiv}$ be a relation on the set $\mathbf{Z}$ of integers, such that $n_1 S_{\equiv} n_2$ exactly if $n_1 - n_2$ is a multiple of 5.

$$\begin{array}{ll} \text{REFLEXIVE:} & nS_{\equiv}n \text{ as } n-n = 0 = 5\times 0 \\ \text{SYMMETRIC:} & n_1-n_2 = 5\times i \;\Rightarrow\; n_2-n_1 = 5\times(-i) \\ \text{TRANSITIVE:} & n_1-n_2 = 5\times i \wedge n_2-n_3 = 5\times j \Rightarrow \\ & n_1-n_3 = 5\times(i+j) \end{array}$$

So this $S_{\equiv}$ is an equivalence relation. □

Theorem 32

If $a_1 \in [a]_{R_{\equiv}}$, $a_2 \in [a]_{R_{\equiv}}$ then $a_1 R_{\equiv} a_2$.

Proof.

Since $a_1 \in [a]_{R_{\equiv}}$, by definition $aR_{\equiv}a_1$, so as $R_{\equiv}$ is symmetric $a_1 R_{\equiv} a$. Also, as $a_2 \in [a]_{R_{\equiv}}$, $aR_{\equiv}a_2$. So since $R_{\equiv}$ is transitive $a_1 R_{\equiv} a_2$. □

Theorem 33

Unequal $R_{\equiv}$ classes are disjoint.

Proof.

We show the contrapositive: classes with a common element are equal. Suppose $a \in [a_1]_{R_{\equiv}}$ and $a \in [a_2]_{R_{\equiv}}$ also. If $b \in [a_1]_{R_{\equiv}}$ then $bR_{\equiv}a$ by Theorem 32. But then as $aR_{\equiv}a_2$, and $R_{\equiv}$ is transitive, $bR_{\equiv}a_2$. That is $b \in [a_2]_{R_{\equiv}}$. So $[a_1]_{R_{\equiv}} \subseteq [a_2]_{R_{\equiv}}$, and by a similar argument $[a_2]_{R_{\equiv}} \subseteq [a_1]_{R_{\equiv}}$. □

Theorem 34

If $R_{\equiv} : A \leftrightarrow A$ is an equivalence, its equivalence classes form a partition of A.

Proof.

DISJOINT: proved in Theorem 33.

NON-EMPTY & COVERING: certainly, since for $\forall a \in A$ we have $a \in [a]_{R_{\equiv}}$. □ □

Consider the relation $S_{\equiv} = \{(n_1, n_2) \mid n_1 - n_2 \text{ is a multiple of } 5\}$ on just the integers $1 \ldots 15$. Its equivalence classes are the subsets in this partition diagram:

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15

ivalence

Example 44

Logical equivalence ≡ is also an equivalence in the relational sense. The set of propositions is infinite, and partitioned by ≡ into infinitely many equivalence classes:

T $p \vee \neg p$ $F \Rightarrow T$ $\ldots$	$\ldots$	$p \vee q$ $\neg p \Rightarrow q$ $\neg q \Rightarrow p$ $\neg(\neg p \wedge \neg q)$ $\ldots$	$\ldots$	F $p \wedge \neg p$ $T \Rightarrow F$ $\ldots$

Closure

If X ∈ {reflexive, symmetric, transitive}, the X closure of a relation R : A ↔ A is the smallest relation on A with property X and R as a subset.

Lemma $A \subseteq C \wedge B \subseteq C \Rightarrow A \cup B \subseteq C$.

- Closure extends a relation just enough to achieve some target property.
- The reflexive closure of a relation contains the identity.
- The symmetric closure of a relation contains the inverse

Theorem 35

The reflexive closure of a relation R on A is $R \cup I_A$, (or $R \cup R^0$).

Proof.

$R \cup I_A$ is indeed reflexive: $I_A \subseteq R \cup I_A$. Also, if R_r is ANY reflexive relation with $R \subseteq R_r$, then since R_r is reflexive $I_A \subseteq R_r$, and by the lemma $R \cup I_A \subseteq R_r$. □

Theorem 36

The symmetric closure of a relation R is $R \cup R^{-1}$.

Proof.

$R \cup R^{-1}$ is indeed symmetric: if $(a_1, a_2) \in R$ then $(a_2, a_1) \in R^{-1}$ so $(a_2, a_1) \in (R \cup R^{-1})$, and similarly for $(a_1, a_2) \in R^{-1}$. Also, if R_s is ANY symmetric relation with $R \subseteq R_s$ then $a_1 R a_2 \Rightarrow a_1 R_s a_2 \Rightarrow a_2 R_s a_1$ so $R^{-1} \subseteq R_s$, and by the lemma $R \cup R^{-1} \subseteq R_s$. □

Proof of theorem 39, $\forall k \geq 1, T^k \subseteq T$ if T is transitive.
By induction on k.
BASE CASE:

$$\begin{aligned} T^1 &= T && \text{by definition} \\ &\subseteq T && \text{as } \subseteq \text{ is reflexive} \end{aligned}$$

INDUCTIVE CASE assuming $T^k \subseteq T$:

$$\begin{aligned} T^{k+1} &= T^k \circ T && \text{by definition} \\ &\subseteq T \circ T && \text{by assumption} \\ &= T^2 && \text{dy definition} \\ &\subseteq T && \text{as } T \text{ is transitive} \end{aligned}$$

□

4/8

Proof of theorem 38, If T is transitive then $T^+ = T$.
We show first $T \subseteq T^+$, then $T^+ \subseteq T$. Recalling the equation $T^+ = T \cup \ldots$ it is immediate that $T \subseteq T^+$. CLAIM: if T is transitive then $\forall k \geq 1, T^k \subseteq T$ — Theorem 39. So T^+ is a union of subsets of T, and $T^+ \subseteq T$. □

5/8

Proof of theorem 37, R's transitive closure is R^+.
R^+ is indeed transitive: if $a_1 R^+ a_2$ and $a_2 R^+ a_3$ there is a path in the digraph of R from a_1 to a_2 and a path from a_2 to a_3, so there is a path from a_1 to a_3. Now if R_t is ANY transitive relation with $R \subseteq R_t$, then as any path in R also exists in R_t we have $R^+ \subseteq R_t^+$. As R_t is transitive, CLAIM: $R_t^+ = R_t$ — see Theorem 38. Therefore, $R^+ \subseteq R_t$. □

Lemma In a digraph with n nodes, if there is any path of > 0 arcs from x to y there must be a path of $\leq n$ arcs.

Definition 67 (R^+)

If $\#A = n$ and $R : A \leftrightarrow A$ then $R^+ = R \cup R^2 \cup \ldots \cup R^n$.

So $a_1 R^+ a_2$ exactly if there is a non-empty path from a_1 to a_2 in the digraph of R.

Theorem 37

R's transitive closure is R^+.

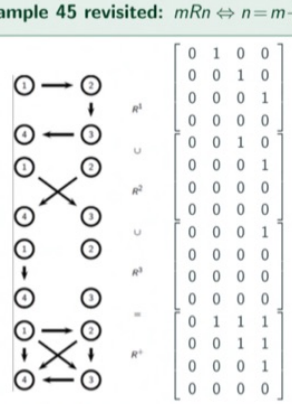
Theorem 38

If T is transitive then $T^+ = T$.

Theorem 39

$\forall k \geq 1, T^k \subseteq T$ if T is transitive.

Example 45 revisited: $mRn \Leftrightarrow n = m+1$



As $R^4 = \emptyset$ it does not contribute to R^+.

- To find a correct closure for a combination of RST properties, establish the properties in alphabetical order.
- Dual-property closures can be obtained by composing closure operations for each property.

Theorem 40

The relation R^* is the reflexive transitive closure of R.

Proof.

R^* is indeed reflexive: $I_A = R^0$ and $R^0 \subseteq R^*$. It is also transi
suppose $a_1 R^* a_2$ and $a_2 R^* a_3$; then if $a_1 = a_2$ or $a_2 = a_3$, it is
immediate that $a_1 R^* a_3$; if not, then $a_1 R^+ a_2$ and $a_2 R^+ a_3$, so
$a_1 R^+ a_3$, hence $a_1 R^* a_3$ since $R^+ \subseteq R^*$. Also, if R_{rt} is ANY
reflexive transitive relation with $R \subseteq R_{rt}$, since R_{rt} is reflexive
$R^0 \subseteq R_{rt}$, and since it is transitive $R^+ \subseteq R_{rt}$; but $R^* = R^0 \cup$
so $R^* \subseteq R_{rt}$. □

Definition 66 extended:
If $X \subseteq \{\text{reflexive}, \text{symmetric}, \text{transitive}\}$ then the X *closure* of R is the smallest relation with ALL the properties in X, and R as a subset.

Definition 68 (R^*)

If R is a relation on a set then $R^* = R^0 \cup R^+$.

Dual-property closures can be obtained by composing closure operations for each property. In one case the order matters.

$$\underset{\substack{\text{reflexive closure}\\ \text{of transitive closure}}}{R^+ \cup R^0} = \underset{\substack{\text{transitive closure of}\\ \text{reflexive closure}}}{(R \cup R^0)^+}$$

$$\underset{\substack{\text{symmetric closure}\\ \text{of reflexive closure}}}{(R \cup R^0) \cup (R \cup R^0)^{-1}} = \underset{\substack{\text{reflexive closure of}\\ \text{symmetric closure}}}{(R \cup R^{-1}) \cup (R \cup R^{-1})^0}$$

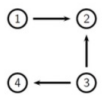
composing closure

BUT

$$\underset{\substack{\text{symmetric closure}\\ \text{of transitive closure}}}{R^+ \cup (R^+)^{-1}} \neq \underset{\substack{\text{transitive closure of}\\ \text{symmetric closure}}}{(R \cup R^{-1})^+}$$

The RHS is the symmetric transitive closure.

Let $A = \{1,2,3,4\}$ and define R on A by the rule $mRn \Leftrightarrow even(n) \wedge (|m-n| = 1)$.

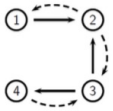


R itself
($= R^+$ as already transitive)

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Example 46

Let $A = \{1,2,3,4\}$ and define R on A by the rule $mRn \Leftrightarrow even(n) \wedge (|m-n| = 1)$.

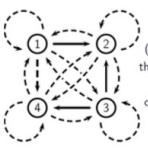


$R^+ \cup (R^+)^{-1}$
(which is NOT transitive)

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Example 46

Let $A = \{1,2,3,4\}$ and define R on A by the rule $mRn \Leftrightarrow even(n) \wedge (|m-n| = 1)$.



$(R \cup R^{-1})^+$
the symmetric transitive closure of R

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$