CS 70 Discrete Mathematics and Probability Theory
Fall 2020

DIS 00

1 Implication

Which of the following implications are always true, regardless of P? Give a counterexample for each false assertion (i.e. come up with a statement $P(x,y)$ that would make the implication false).

(a) $\forall x \forall y P(x,y) \implies \forall y \forall x P(x,y)$.

(b) $\forall x \exists y P(x,y) \implies \exists y \forall x P(x,y)$.

(c) $\exists x \forall y P(x,y) \implies \forall y \exists x P(x,y)$.

Solution:

(a) True. For all can be switched if they are adjacent; since $\forall x, \forall y$ and $\forall y, \forall x$ means for all x and y in our universe.

(b) False. Let $P(x,y)$ be $x < y$, and the universe for x and y be the integers. Or let $P(x,y)$ be $x = y$ and the universe be any set with at least two elements. In both cases, the antecedent is true and the consequence is false, thus the entire implication statement is false.

(c) True. The first statement says that there is an x, say x' where for every y, $P(x,y)$ is true. Thus, one can choose $x = x'$ for the second statement and that statement will be true again for every y. Note: 4c and 4d are not logically equivalent. In fact, the converse of 4d is 4c, which we saw is false.

2 Equivalences with Quantifiers

Evaluate whether the expressions on the left and right sides are equivalent in each part, and briefly justify your answers.

(a)	$\forall x \left((\exists y\, Q(x,y)) \Rightarrow P(x)\right)$	$\forall x \exists y \left(Q(x,y) \Rightarrow P(x)\right)$
(b)	$\neg \exists x \forall y \left(P(x,y) \Rightarrow \neg Q(x,y)\right)$	$\forall x \left((\exists y\, P(x,y)) \wedge (\exists y\, Q(x,y))\right)$
(c)	$\forall x \exists y \left(P(x) \Rightarrow Q(x,y)\right)$	$\forall x \left(P(x) \Rightarrow (\exists y\, Q(x,y))\right)$

Solution:

(a) Not equivalent.

Justification: We can rewrite the left side as as $\forall x \left((\neg(\exists y\, Q(x,y))) \vee P(x)\right)$ and the right side as $\forall x\, \exists y \left(\neg Q(x,y) \vee P(x)\right)$ Applying the negation on the left side of the equivalence

$(\neg(\exists y Q(x,y)))$ changes the $\exists y$ to $\forall y$, and the two sides are clearly not the same. Another approach to the problem is to consider by linguistic example. Let x and y span the universe of all people, and let $Q(x,y)$ mean "Person x is Person y's offspring", and let $P(x)$ mean "Person x likes tofu". The right side claims that, for all Persons x, there exists some Person y such that either Person x is not Person y's offspring or that Person x likes tofu. The left side claims that, for all Persons x, if there exists a parent of Person x, then Person x likes tofu. Obviously, these are not the same.

(b) Not equivalent.

Justification: Using De Morgan's Law to distribute the negation on the left side yields

$$\forall x\, \exists y\, \big(P(x,y) \wedge Q(x,y)\big).$$

But $\exists$ does not distribute over $\wedge$. There could exist different values of y such that $P(x,y)$ and $Q(x,y)$ for a given x, but not necessarily the same value.

(c) Equivalent.

Justification: We can rewrite the left side as $\forall x\, \exists y\, \big(\neg P(x) \vee Q(x,y)\big)$ and the right side as $\forall x\, \big(\neg P(x) \vee (\exists y\, Q(x,y))\big)$. Clearly, the two sides are the same if $\neg P(x)$ is true. If $\neg P(x)$ is false, then the two sides are still the same, because $\forall x\, \exists y\, \big(\text{False} \vee Q(x,y)\big) \equiv \forall x\, \big(\text{False} \vee (\exists y\, Q(x,y))\big)$.

3 XOR

The truth table of XOR (denoted by $\oplus$) is as follows.

A	B	A $\oplus$ B
F	F	F
F	T	T
T	F	T
T	T	F

1. Express XOR using only $(\wedge, \vee, \neg)$ and parentheses.
2. Does $(A \oplus B)$ imply $(A \vee B)$? Explain briefly.
3. Does $(A \vee B)$ imply $(A \oplus B)$? Explain briefly.

Solution:

1. These are all correct:
   - $A \oplus B = (A \wedge \neg B) \vee (\neg A \wedge B)$
     Notice that there are only two instances when $A \oplus B$ is true: (1) when A is true and B is false, or (2) when B is true and A is false. The clause $(A \wedge \neg B)$ is only true when (1) is, and the clause $(\neg A \wedge B)$ is only true when (2) is.

- $A \oplus B = (A \vee B) \wedge (\neg A \vee \neg B)$
  Another way to think about XOR is that exactly one of A and B needs to be true. This also means exactly one of $\neg A$ and $\neg B$ needs to be true. The clause $(A \vee B)$ tells us *at least* one of A and B needs to be true. In order to ensure that one of A or B is also false, we need the clause $(\neg A \vee \neg B)$ to be satisfied as well.
- $A \oplus B = (A \vee B) \wedge \neg(A \wedge B)$
  This is the same as the previous, with De Morgan's law applied to equate $(\neg A \vee \neg B)$ to $\neg(A \wedge B)$.

2. Yes. $(A \oplus B) \implies (A \wedge \neg B) \vee (\neg A \wedge B) \implies (A \vee B)$. When $(A \oplus B)$ is true, at least one of A or B is true, which makes $(A \vee B)$ true as well.

3. No. When A and B are both true, then $(A \vee B)$ is true, but $(A \oplus B)$ is false.

4 Truth Tables

Determine whether the following equivalences hold, by writing out truth tables. Clearly state whether or not each pair is equivalent.

(a) $P \wedge (Q \vee P) \equiv P \wedge Q$

(b) $(P \vee Q) \wedge R \equiv (P \wedge R) \vee (Q \wedge R)$

(c) $(P \wedge Q) \vee R \equiv (P \vee R) \wedge (Q \vee R)$

Solution:

(a) Not equivalent.

P	Q	$P \wedge (Q \vee P)$	$P \wedge Q$
T	T	T	T
T	F	T	F
F	T	F	F
F	F	F	F

(b) Equivalent.

P	Q	R	$(P \vee Q) \wedge R$	$(P \wedge R) \vee (Q \wedge R)$
T	T	T	T	T
T	T	F	F	F
T	F	T	T	T
T	F	F	F	F
F	T	T	T	T
F	T	F	F	F
F	F	T	F	F
F	F	F	F	F

(c) Equivalent.

P	Q	R	$(P \wedge Q) \vee R$	$(P \vee R) \wedge (Q \vee R)$
T	T	T	T	T
T	T	F	T	T
T	F	T	T	T
T	F	F	F	F
F	T	T	T	T
F	T	F	F	F
F	F	T	T	T
F	F	F	F	F

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 01A

1 Proof Practice

(a) Prove that $\forall n \in \mathbb{N}$, if n is odd, then n^2+1 is even. (Recall that n is odd if $n = 2k+1$ for some natural number k.)

(b) Prove that $\forall x, y \in \mathbb{R}$, $\min(x,y) = (x+y-|x-y|)/2$. (Recall, that the definition of absolute value for a real number z, is

$$|z| = \begin{cases} z, & z \geq 0 \\ -z, & z < 0 \end{cases}$$

(c) Suppose $A \subseteq B$. Prove $\mathscr{P}(A) \subseteq \mathscr{P}(B)$. (Recall that $A' \in \mathscr{P}(A)$ if and only if $A' \subseteq A$.)

Solution:

(a) We will use a direct proof. Assume n is odd. By the definition of odd numbers, $n = 2k+1$ for some natural number k. Substituting into the expression n^2+1, we get $(2k+1)^2+1$. Simplifying the expression yields $4k^2+4k+2$. This can be rewritten as $2 \times (2k^2+2k+1)$. Since $2k^2+2k+1$ is a natural number, by the definition of even numbers, n^2+1 is even.

(b) We will use a proof by cases. Again, the definition of the absolute value function for real number z is

$$|z| = \begin{cases} z, & z \geq 0 \\ -z, & z < 0 \end{cases}$$

Case 1: $x < y$. This means $|x-y| = y-x$. Substituting this into the formula on the right hand side, we get

$$\frac{x+y-y+x}{2} = x = \min(x,y).$$

Case 2: $x \geq y$. This means $|x-y| = x-y$. Substituting this into the formula on the right hand side, we get

$$\frac{x+y-x+y}{2} = y = \min(x,y).$$

(c) Suppose $A' \in \mathscr{P}(A)$, that is, $A' \subseteq A$ (by the definition of the power set). We must prove that for any such A', we also have that $A' \in \mathscr{P}(B)$, that is, $A' \subseteq B$.

Let $x \in A'$. Then, since $A' \subseteq A$, $x \in A$. Since $A \subseteq B$, $x \in B$. We have shown $(\forall x \in A')\, x \in B$, so $A' \subseteq B$.

Since the previous argument works for any $A' \subseteq A$, we have proven $(\forall A' \in \mathscr{P}(A))\, A' \in \mathscr{P}(B)$. So, we conclude $\mathscr{P}(A) \subseteq \mathscr{P}(B)$ as desired.

2 Preserving Set Operations

For a function f, define the image of a set X to be the set $f(X) = \{y \mid y = f(x) \text{ for some } x \in X\}$. Define the inverse image or preimage of a set Y to be the set $f^{-1}(Y) = \{x \mid f(x) \in Y\}$. Prove the following statements, in which A and B are sets. By doing so, you will show that inverse images preserve set operations, but images typically do not.

Recall: For sets X and Y, $X = Y$ if and only if $X \subseteq Y$ and $Y \subseteq X$. To prove that $X \subseteq Y$, it is sufficient to show that $(\forall x)\ ((x \in X) \implies (x \in Y))$.

(a) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

(b) $f(A \cup B) = f(A) \cup f(B)$.

Solution:

In order to prove equality $A = B$, we need to prove that A is a subset of B, $A \subseteq B$ and that B is a subset of A, $B \subseteq A$. To prove that LHS is a subset of RHS we need to prove that if an element is a member of LHS then it is also an element of the RHS.

(a) Suppose $x \in f^{-1}(A \cup B)$ which means that $f(x) \in A \cup B$. Then either $f(x) \in A$, in which case $x \in f^{-1}(A)$, or $f(x) \in B$, in which case $x \in f^{-1}(B)$, so in either case we have $x \in f^{-1}(A) \cup f^{-1}(B)$. This proves that $f^{-1}(A \cup B) \subseteq f^{-1}(A) \cup f^{-1}(B)$.

Now, suppose that $x \in f^{-1}(A) \cup f^{-1}(B)$. Suppose, without loss of generality, that $x \in f^{-1}(A)$. Then $f(x) \in A$, so $f(x) \in A \cup B$, so $x \in f^{-1}(A \cup B)$. The argument for $x \in f^{-1}(B)$ is the same. Hence, $f^{-1}(A) \cup f^{-1}(B) \subseteq f^{-1}(A \cup B)$.

(b) Suppose that $x \in A \cup B$. Then either $x \in A$, in which case $f(x) \in f(A)$, or $x \in B$, in which case $f(x) \in f(B)$. In either case, $f(x) \in f(A) \cup f(B)$, so $f(A \cup B) \subseteq f(A) \cup f(B)$.

Now, suppose that $y \in f(A) \cup f(B)$. Then either $y \in f(A)$ or $y \in f(B)$. In the first case, there is an element $x \in A$ with $f(x) = y$; in the second case, there is an element $x \in B$ with $f(x) = y$. In either case, there is an element $x \in A \cup B$ with $f(x) = y$, which means that $y \in f(A \cup B)$. So $f(A) \cup f(B) \subseteq f(A \cup B)$.

The purpose of this problem is to gain familiarity to naming thing precisely. In particular, we named an element in the LHS (or the pre-image of the LHS) and then argued about whether that element or its image was in the right hand side. By explicitly naming an element generically where it could be *any element in the set, we could argue about its membership in a set and or its image or preimage. With these different concepts floating around it is helpful to be clear in the argument.*

3 Fermat's Contradiction

Prove that $2^{1/n}$ is not rational for any integer $n \geq 3$. (*Hint*: Use Fermat's Last Theorem. It states that there exists no positive integers a, b, c s.t. $a^n + b^n = c^n$ for $n \geq 3$.)

Solution:

If not, then there exists an integer $n \geq 3$ such that $2^{1/n} = \frac{p}{q}$ where p, q are positive integers. Thus, $2q^n = p^n$, and this implies

$$q^n + q^n = p^n,$$

which is a contradiction to the Fermat's Last Theorem.

4 Pebbles

Suppose you have a rectangular array of pebbles, where each pebble is either red or blue. Suppose that for every way of choosing one pebble from each column, there exists a red pebble among the chosen ones. Prove that there must exist an all-red column.

Solution: We give a proof by contraposition. Suppose there does not exist an all-red column. This means that, in each column, we can find a blue pebble. Therefore, if we take one blue pebble from each column, we have a way of choosing one pebble from each column without any red pebbles. This is the negation of the original hypothesis, so we are done.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 01B

1 Induction

Prove the following using induction:

(a) Let a and b be integers with $a \neq b$. For all natural numbers $n \geq 1$, $(a^n - b^n)$ is divisible by $(a-b)$.

(b) For all natural numbers n, $(2n)! \leq 2^{2n}(n!)^2$. [Note that 0! is defined to be 1.]

Solution:

(a)
- Base case (n=1): $P(1)$ says that $(a^1 - b^1) = (a-b)$, which is trivially divisible by $(a-b)$.
- Inductive Hypothesis: For arbitrary $n = k \geq 1$, assume that $P(k)$ is true: $a^k - b^k = q_k(a-b), \quad q_k \in \mathbb{Z}$.
- Inductive Step: Prove the statement for $n = k+1$: $P(k+1)$ gives $a^{k+1} - b^{k+1} = q_{k+1}(a-b), \quad q_{k+1} \in \mathbb{Z}$.
  The goal is to express $(a^{k+1} - b^{k+1})$ in terms of $(a^k - b^k)$ and $(a-b)$ (since both of these are divisible by $(a-b)$ which we know their summation is also divisible by $(a-b)$). We can do this as follows:

$$\begin{aligned} a^{k+1} - b^{k+1} &= a(a^k - b^k) + b^k(a-b) \\ &= a \underbrace{(a^k - b^k)}_{\text{divisble by } (a-b)} \qquad \text{(Inductive Hypothesis)} \\ &\quad + b^k(a-b) \end{aligned}$$

Alternatively we can write the following:

$$a^{k+1} - b^{k+1} = \frac{1}{2}[(a+b) \underbrace{(a^k - b^k)}_{\text{divisble by } (a-b)} + (a-b)(a^k + b^k)]$$

again each of the terms in the parentheses is divisible by $(a-b)$. For this argument, one should really also verify that both of those terms are even (so that both are integers when divided by 2); but that's easy to see since the parity of the two terms in the two products is equal.

Thus, $a^{k+1} - b^{k+1} = q_{k+1}(a-b), \quad q_{k+1} \in \mathbb{Z}$.

Hence, $(a^n - b^n)$ is divisble by $(a-b)$ for all $n \geq 1$ by induction.

(b)

- Base case (n=0): $P(0)$ asserts that $(2(0))! = 1 = 2^{(2(0))}(0!)^2$. So we showed the base case is correct.
- Inductive Hypothesis: For arbitrary $n = k \geq 0$, assume that $P(k)$ is correct which leads to $(2k)! \leq 2^{2k}(k!)^2$.
- Inductive Step: Prove the statement for $n = k+1$: i.e., prove that $(2(k+1))! \leq 2^{2(k+1)}((k+1)!)^2$.

$$\begin{aligned}
(2(k+1))! &= (2k)!(2k+2)(2k+1) \\
&\leq 2^{2k}(k!)^2 2(k+1)(2k+1) \qquad \text{(Inductive Hypothesis)} \\
&= 2^{2k+1}(k+1)!k!(2k+1) \\
&\leq 2^{2k+1}(k+1)!k!(2k+2) \\
&\leq 2^{2(k+1)}(k+1)!(k+1)! \\
&= 2^{2(k+1)}((k+1)!)^2.
\end{aligned}$$

Thus, $(2(k+1))! \leq 2^{2(k+1)}((k+1)!)^2$.
Hence, $(2n)! \leq 2^{2n}(n!)^2$ holds for all $n \geq 0$ by induction.

2 Make It Stronger

Suppose that the sequence $a_1, a_2, \ldots$ is defined by $a_1 = 1$ and $a_{n+1} = 3a_n^2$ for $n \geq 1$. We want to prove that

$$a_n \leq 3^{2^n}$$

for every positive integer n.

(a) Suppose that we want to prove this statement using induction, can we let our induction hypothesis be simply $a_n \leq 3^{2^n}$? Show why this does not work.

(b) Try to instead prove the statement $a_n \leq 3^{2^n-1}$ using induction. Does this statement imply what you tried to prove in the previous part?

Solution:

(a) Try to prove that for every $n \geq 1$, we have $a_n \leq 3^{2^n}$ by induction.

Base Case: For $n = 1$ we have $a_1 = 1 \leq 3^{2^1} = 9$.
Inductive Step: For some $n \geq 1$, we assume $a_n \leq 3^{2^n}$. Now, consider $n+1$. We can write:

$$a_{n+1} = 3a_n^2 \leq 3(3^{2^n})^2 = 3 \times 3^{2 \times 2^n} = 3 \times 3^{2^{n+1}} = 3^{2^{n+1}+1}.$$

However, what we wanted was to get an inequality of the form: $a_{n+1} \leq 3^{2^{n+1}}$. There is an extra $+1$ in the exponent of what we derived.

(b) This time the induction works.

Base Case: For $n = 1$ we have $a_1 = 1 \leq 3^{2-1} = 3$.
Inductive Step: For some $n \geq 1$ we assume $a_n \leq 3^{2^n-1}$. Now, consider $n+1$. We can write:

$$a_{n+1} = 3a_n^2 \leq 3 \times (3^{2^n-1})^2 = 3 \times 3^{2\times(2^n-1)} = 3 \times 3^{2^{n+1}-2} = 3^{2^{n+1}-1}.$$

This is exactly the induction hypothesis for $n+1$. Note that for every $n \geq 1$, we have $2^n - 1 \leq 2^n$ and therefore $3^{2^n-1} \leq 3^{2^n}$. This means that our modified hypothesis which we proved here does indeed imply what we wanted to prove in the previous part. This is called "strengthening" the induction hypothesis because we proved a stronger statement and by proving that statement to be true, we proved our original statement to be true as well.

3 Binary Numbers

Prove that every positive integer n can be written in binary. In other words, prove that we can write

$$n = c_k \cdot 2^k + c_{k-1} \cdot 2^{k-1} + \cdots + c_1 \cdot 2^1 + c_0 \cdot 2^0,$$

where $k \in \mathbb{N}$ and $c_k \in \{0, 1\}$.

Solution:

Prove by strong induction on n. (Note that this is the first discussion where the students use strong induction, so it is important that this problem be done in an interactive way that shows them how simple induction gets stuck.)

The key insight here is that if n is divisible by 2, then it is easy to get a bit string representation of $(n+1)$ from that of n. However, if n is not divisible by 2, then $(n+1)$ will be, and its binary representation will be more easily derived from that of $(n+1)/2$. More formally:

- Base Case: $n = 1$ can be written as 1×2^0.
- Inductive Step: Assume that the statement is true for all $1 \leq m \leq n$, where n is arbitrary. Now, we need to consider $n+1$. If $n+1$ is divisible by 2, then we can apply our inductive hypothesis to $(n+1)/2$ and use its representation to express $n+1$ in the desired form.

$$(n+1)/2 = c_k \cdot 2^k + c_{k-1} \cdot 2^{k-1} + \cdots + c_1 \cdot 2^1 + c_0 \cdot 2^0$$
$$n+1 = 2 \cdot (n+1)/2 = c_k \cdot 2^{k+1} + c_{k-1} \cdot 2^k + \cdots + c_1 \cdot 2^2 + c_0 \cdot 2^1 + 0 \cdot 2^0.$$

Otherwise, n must be divisible by 2 and thus have $c_0 = 0$. We can obtain the representation of $n+1$ from n as follows:

$$n = c_k \cdot 2^k + c_{k-1} \cdot 2^{k-1} + \cdots + c_1 \cdot 2^1 + 0 \cdot 2^0$$
$$n+1 = c_k \cdot 2^k + c_{k-1} \cdot 2^{k-1} + \cdots + c_1 \cdot 2^1 + 1 \cdot 2^0$$

Therefore, the statement is true.

Note: In proofs using simple induction, we only use $P(n)$ in order to prove $P(n+1)$. Simple induction gets stuck here because in order to prove $P(n+1)$ in the inductive step, we need to assume more than just $P(n)$. This is because it is not immediately clear how to get a representation for $P(n+1)$ using just $P(n)$, particularly in the case that $n+1$ is divisible by 2. As a result, we assume the statement to be true for all of $1, 2, \ldots, n$ in order to prove it for $P(n+1)$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020

DIS 02A

1 Stable Matching

Consider the set of candidates $C = \{1, 2, 3\}$ and the set of jobs $J = \{A, B, C\}$ with the following preferences.

C	J		
1	A	B	C
2	B	A	C
3	A	B	C

J	C		
A	2	1	3
B	1	2	3
C	1	2	3

Run the applicant propose-and-reject algorithm on this example. How many days does it take and what is the resulting pairing? (Show your work)

Solution:

The algorithm takes 3 days to produce a matching. The resulting pairing is {(A, 1), (B, 2), (C, 3)}

Jobs	Day 1	Day 2	Day 3
A	①,3	①	①
B	②	②,3	②
C			③

2 Good, Better, Best

In a particular instance of the stable marriage problem with n applicants and n jobs, it turns out that there are exactly three distinct stable matchings, S_1, S_2, and S_3. Also, each applicant m has a different partner in the three matchings. Therefore each applicant has a clear preference ordering of the three matchings (according to the ranking of his partners in his preference list). Now, suppose for applicant m_1, this order is $S_1 > S_2 > S_3$.

Prove that every applicant has the same preference ordering $S_1 > S_2 > S_3$.

Hint: Recall that a applicant-optimal matching always exists and can be generated using applicant proposes matching algorithm. By reversing the roles of stable matching algorithm, what other matching can we generate?

Solution:

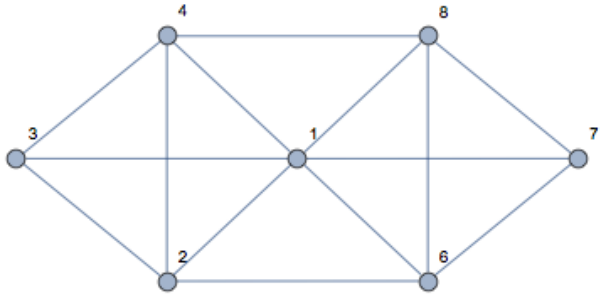
In class, you were given the traditional propose-and-reject algorithm, which was guaranteed to produce a applicant-optimal matching. By switching applicant's and jobs's roles, you would be

guaranteed to produce a job-optimal matching, which, by a lemma from class, would also be applicant-pessimal. By the very fact that these algorithms exist and have been proven to work in this way, you're guaranteed that an applicant-optimal and a applicant-pessimal matching always exist.

Since there are only three matchings in this particular stable matching instance, we thus know that one of them must be applicant-optimal and one must be applicant-pessimal. Since m_1 prefers S_1 above the other stable matchings, only that one can be applicant-optimal by definition of applicant-optimality. Similarly, since m_1 prefers S_3 the least, it must be the applicant-pessimal. Therefore, again from definitions of optimality/pessimality, since each applicant has different matches in the three stable matchings, they *must* strictly prefer S_1 to both of the others, and they *must* like S_3 strictly less than both of the others. Thus, each applicant's preference order of stable matchings must be S_1, S_2, S_3.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 02B

1 Eulerian Tour and Eulerian Walk



(a) Is there an Eulerian tour in the graph above? If no, give justification. If yes, provide an example.

(b) Is there an Eulerian walk in the graph above? An Eulerian walk is a walk that uses each edge exactly once. If no, give justification. If yes, provide an example.

(c) What is the condition that there is an Eulerian walk in an undirected graph? Briefly justify your answer.

Solution:

(a) No. Two vertices have odd degree.

(b) Yes. One of the two vertices with odd degree must be the starting vertex, and the other one must be the ending vertex. For example: $3,4,2,1,3,2,6,1,4,8,1,7,8,6,7$ will be an Eulerian walk (the numbers are the vertices visited in order). Note that there are 14 edges in the graph.

(c) This solution is long and in depth. Please read slowly, and don't worry if it takes multiple read-throughs since this is dense mathematical text.

An undirected graph has an Eulerian walk if and only if it is connected (except for isolated vertices) and has at most two odd degree vertices. Note that there is no graph with only one odd degree vertex (this is a result of the Handshake lemma, which we will prove in the next question). An Eulerian tour is also an Eulerian walk which starts and ends at the same vertex. We have already seen in the lectures, that an undirected graph G has an Eulerian tour if and only if G is connected (except for isolated vertices) and all its vertices have even degree. We will now prove that a graph G has an Eulerian walk with distinct starting and ending vertex, if and only if it is connected (except for isolated vertices) and has exactly two odd degree vertices.

Justifications: *Only if.* Suppose there exists an Eulerian walk, say starting at u and ending at v (note that u and v are distinct). Then all the vertices that lie on this walk are connected to each other and all the vertices that do not lie on this walk (if any) must be isolated. Thus the graph is connected (except for isolated vertices). Moreover, every intermediate visit to a vertex in this walk is being paired with two edges, and therefore, except for u and v, all other vertices must be of even degree.

If. First, note that for a connected graph with no odd degree vertices, we have shown in the lectures that there is an Eulerian tour, which implies an Eulerian walk. Thus, let us consider the case of two odd degree vertices.

Solution 1: Take the two odd degree vertices u and v, and add a vertex w with two edges (u,w) and (w,v). The resulting graph G' has only vertices of even degree (we added one to the degree of u and v and introduced a vertex of degree 2) and is still connected. So, we can find an Eulerian tour on G'. Now, delete the component of the tour that uses edges (u,w) and (w,v). The part of the tour that is left is now an Eulerian walk from u to v on the original graph, since it traverses every edge on the original graph.

Solution 2: Alternatively, we can construct an algorithm quite similar to the FindTour algorithm with splicing described in the graphs note.

Suppose G is connected (except for isolated vertices) and has exactly two odd degree vertices, say u and v. First remove the isolated vertices if any. Since u and v belong to a connected component, one can find a path from u to v. Consider the graph obtained by removing the edges of the path from the graph. In the resulting graph, all the vertices have even degree. Hence, for each connected component of the residual graph, we find an Eulerian tour. (Note that the graph obtained by removing the edges of the path can be disconnected.) Observe that an Eulerian walk is simply an edge-disjoint walk that covers all the edges. What we just did is decomposing all the edges into a path from u to v and a bunch of edge-disjoint Eulerian tours. A path is clearly an edge-disjoint walk. Then, given an edge-disjoint walk and an edge-disjoint tour such that they share at least one common vertex, one can combine them into an edge-disjoint walk simply by augmenting the walk with the tour at the common vertex. Therefore we can combine all the edge-disjoint Eulerian tours into the path from u to v to make up an Eulerian walk from u to v.

2 Banquet Arrangement

In the words of the great Ana Lynch, "Let's have a kiki."

Suppose n people are attending a kiki, and each of them has at least m friends ($2 \leq m \leq n$), where friendship is mutual. Prove that we can put at least $m+1$ of the attendants on the same round table, so that each person sits next to his or her friends on both sides.

Solution: Let each person be a vertex and add an edge between two people if they are friends. Thus we have a graph with n vertices. Since each of them has at least m friends, we know that all the vertices in the graph have degree at least m. Suppose we find a cycle of length at least $m+1$ in

this graph, say $C = \{v_0, v_1, \ldots, v_k\}$, where $k \geq m$. If we place these $k+1$ people at the round table in the order given by the cycle C, they observe that each person sits next to his or her friends since he/she has an edge with him/her in the corresponding graph. Thus we can rephrase the problem in graph theory terms as follows: given that all the vertices in an n-vertex graph have degree at least m, show that there exists a cycle containing at least $m+1$ vertices.

Let $P = v_0 v_1 \ldots v_l$ be a longest path in the graph. Such a path exists because the length of paths is bounded above by n. All neighbors of v_0 must be in P, since otherwise P can be extended to be even longer by appending this edge at the beginning of path P. Let k be the maximum index of neighbors of v_0 along P. Since v_0 has at least m neighbors, we must have $k \geq m$. Then $v_0 v_1 \ldots v_k v_0$ gives us the desired cycle.

3 Not everything is normal: Odd-Degree Vertices

Claim: Let $G = (V, E)$ be an undirected graph. The number of vertices of G that have odd degree is even.

Prove the claim above using:

(i) Direct proof (e.g., counting the number of edges in G). *Hint: in lecture, we proved that* $\sum_{v \in V} \deg v = 2|E|$.

(ii) Induction on $m = |E|$ (number of edges)

(iii) Induction on $n = |V|$ (number of vertices)

Solution:

Let $V_{\text{odd}}(G)$ denote the set of vertices in G that have odd degree. We prove that $|V_{\text{odd}}(G)|$ is even.

(i) Let d_v denote the degree of vertex v (so $d_v = |N_v|$, where N_v is the set of neighbors of v). Observe that

$$\sum_{v \in V} d_v = 2m$$

because every edge is counted exactly twice when we sum the degrees of all the vertices. Now partition V into the odd degree vertices $V_{\text{odd}}(G)$ and the even degree vertices $V_{\text{odd}}(G)^c$, so we can write

$$\sum_{v \in V_{\text{odd}}(G)} d_v = 2m - \sum_{v \notin V_{\text{odd}}(G)} d_v.$$

Both terms in the right-hand side above are even ($2m$ is even, and each term d_v is even because we are summing over even degree vertices $v \notin V_{\text{odd}}(G)$). So for the left-hand side $\sum_{v \in V_{\text{odd}}(G)} d_v$ to be even, we must have an even number of terms, since each term in the summation is odd. Therefore, there must be an even number of odd-degree vertices, namely, $|V_{\text{odd}}(G)|$ is even.

(ii) We use induction on $m \geq 0$.

Base case $m = 0$: If there are no edges in G, then all vertices have degree 0, so $V_{\text{odd}}(G) = \varnothing$.

Inductive hypothesis: Assume $|V_{\text{odd}}(G)|$ is even for all graphs G with m edges.

Inductive step: Let G be a graph with $m+1$ edges. Remove an arbitrary edge $\{u,v\}$ from G, so the resulting graph G' has m edges. By the inductive hypothesis, we know $|V_{\text{odd}}(G')|$ is even. Now add the edge $\{u,v\}$ to get back the original graph G. Note that u has one more edge in G than it does in G', so $u \in V_{\text{odd}}(G)$ if and only if $u \notin V_{\text{odd}}(G')$. Similarly, $v \in V_{\text{odd}}(G)$ if and only if $v \notin V_{\text{odd}}(G')$. The degrees of all other vertices are unchanged in going from G' to G. Therefore,

$$V_{\text{odd}}(G) = \begin{cases} V_{\text{odd}}(G') \cup \{u,v\} & \text{if } u,v \notin V_{\text{odd}}(G') \\ V_{\text{odd}}(G') \setminus \{u,v\} & \text{if } u,v \in V_{\text{odd}}(G') \\ (V_{\text{odd}}(G') \setminus \{u\}) \cup \{v\} & \text{if } u \in V_{\text{odd}}(G'), v \notin V_{\text{odd}}(G') \\ (V_{\text{odd}}(G') \setminus \{v\}) \cup \{u\} & \text{if } u \notin V_{\text{odd}}(G'), v \in V_{\text{odd}}(G') \end{cases}$$

so we see that $|V_{\text{odd}}(G)| - |V_{\text{odd}}(G')| \in \{-2,0,2\}$. Since $|V_{\text{odd}}(G')|$ is even, we conclude $|V_{\text{odd}}(G)|$ is also even.

(iii) We use induction on $n \geq 1$.

Base case $n = 1$: If G only has 1 vertex, then that vertex has degree 0, so $V_{\text{odd}}(G) = \varnothing$.

Inductive hypothesis: Assume $|V_{\text{odd}}(G)|$ is even for all graphs G with n vertices.

Inductive step: Let G be a graph with $n+1$ vertices. Remove a vertex v and all edges adjacent to it from G. The resulting graph G' has n vertices, so by the inductive hypothesis, $|V_{\text{odd}}(G')|$ is even. Now add the vertex v and all edges adjacent to it to get back the original graph G. Let $N_v \subseteq V$ denote the neighbors of v (i.e., all vertices adjacent to v). Among the neighbors N_v, the vertices in the intersection $A = N_v \cap V_{\text{odd}}(G')$ had odd degree in G', so they now have even degree in G. On the other hand, the vertices in $B = N_v \cap V_{\text{odd}}(G')^c$ had even degree in G', and they now have odd degree in G. The vertex v itself has degree $|N_v|$, so $v \in V_{\text{odd}}(G)$ if and only if $|N_v|$ is odd. We now consider two cases:

(a) Suppose $|N_v|$ is even, so $v \notin V_{\text{odd}}(G)$. Then

$$V_{\text{odd}}(G) = (V_{\text{odd}}(G') \setminus A) \cup B$$

so $|V_{\text{odd}}(G)| = |V_{\text{odd}}(G')| - |A| + |B|$. Note that A and B are disjoint and their union equals N_v, so $|A| + |B| = |N_v|$. Therefore, we can write $|V_{\text{odd}}(G)|$ as

$$|V_{\text{odd}}(G)| = |V_{\text{odd}}(G')| + |N_v| - 2|A|$$

which is even, since $|V_{\text{odd}}(G')|$ is even by the inductive hypothesis, and $|N_v|$ is even by assumption.

(b) Suppose $|N_v|$ is odd, so $v \in V_{\text{odd}}(G)$. Then

$$V_{\text{odd}}(G) = (V_{\text{odd}}(G') \setminus A) \cup B \cup \{v\}$$

so, again using the relation $|A| + |B| = |N_v|$, we can write

$$|V_{\text{odd}}(G)| = |V_{\text{odd}}(G')| - |A| + |B| + 1 = |V_{\text{odd}}(G')| + (|N_v| + 1) - 2|A|$$

which is even, since $|V_{\text{odd}}(G')|$ is even by the inductive hypothesis, and $|N_v|$ is odd by assumption.

This completes the inductive step and the proof.

Note how this proof is more complicated than the proof in part (ii), even though they are both using induction. This tells you that choosing the right variable to induct on can simplify the proof.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 03A

1 Cube Dual

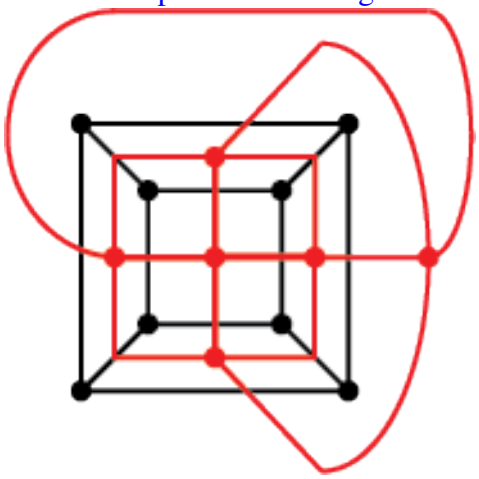
We define a graph G by letting the vertices be the corners of a cube and having edges connecting adjacent corners. Define the *dual* of a planar graph G to be a graph G', constructed by replacing each face in G with a vertex, and an edge between every vertex in G' if the respective faces are adjacent in G.

(a) Draw a planar representation of G and the corresponding dual graph. Is the dual graph planar? (Hint: think about the act of drawing the dual)

(b) Is G' bipartite?

Solution:

(a) Here is one possible drawing of the cube (in black) with its dual (in red):



As seen in the drawing, the dual is indeed planar.

(b) From the drawing, G' is not bipartite. This is a reminder that connecting the middle of every face on a cube does not result in another cube, which would be bipartite!

2 True or False

(a) Any pair of vertices in a tree are connected by exactly one path.

(b) Adding an edge between two vertices of a tree creates a new cycle.

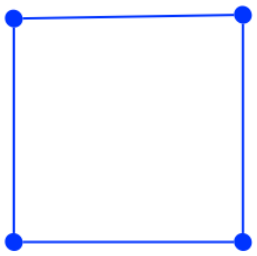
(c) Adding an edge in a connected graph creates exactly one new cycle.

Solution:

(a) **True.**
Pick any pair of vertices x, y. We know there is a path between them since the graph is connected. We will prove that this path is unique by contradiction: Suppose there are two distinct paths from x to y. At some point (say at vertex a) the paths must diverge, and at some point (say at vertex b) they must reconnect. So by following the first path from a to b and the second path in reverse from b to a we get a cycle. This gives the necessary contradiction.

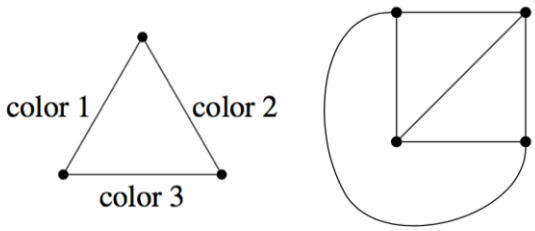
(b) **True.**
Pick any pair of vertices x, y not connected by an edge. We prove that adding the edge $\{x,y\}$ will create a cycle. From part (a), we know that there is a unique path between x and y. Therefore, adding the edge $\{x,y\}$ creates a cycle obtained by following the path from x to y, then following the edge $\{x,y\}$ from y back to x.

(c) **False.**
In the following graph adding an edge creates two cycles.



3 Edge Colorings

An edge coloring of a graph is an assignment of colors to edges in a graph where any two edges incident to the same vertex have different colors. An example is shown on the left.



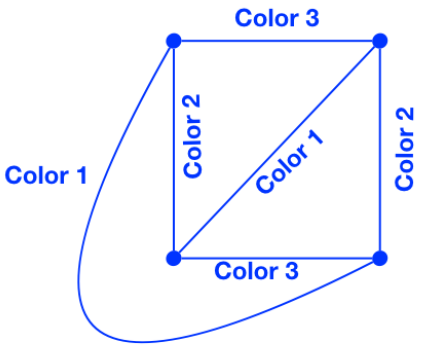
(a) Show that the 4 vertex complete graph above can be 3 edge colored. (Use the numbers $1,2,3$ for colors. A figure is shown on the right.)

(b) Prove that any graph with maximum degree $d \geq 1$ can be edge colored with $2d-1$ colors.

(c) Show that a tree can be edge colored with d colors where d is the maximum degree of any vertex.

Solution:

(a) Three color a triangle. Now add the fourth vertex notice, call it vertex u. For any edge, say $\{u,v\}$ from this fourth vertex u, observe that the vertex v has two edges from before and hence there a third color available for the edge $\{u,v\}$.



(b) We will use induction on the number of edges n in the graph to prove the statement: If a graph G has $n \geq 0$ edges and the maximum degree of any vertex is d, then G can be colored with $2d-1$ colors.
Base case (n = 0). If there are no edges in the graph, then there is nothing to be colored and the statement holds trivially.
Inductive hypothesis. Suppose for $n = k \geq 0$, the statement holds.
Inductive step. Consider a graph G with $n = k+1$ edges. Remove an edge of your choice, say e from G. Note that in the resulting graph the maximum degree of any vertes is $d' \leq d$. By the inductive hypothesis, we can color this graph using $2d'-1$ colors and hence with $2d-1$ colors too. The removed edge is incident to two vertices each of which is incident to at most $d-1$ other edges, and thus at most $2(d-1) = 2d-2$ colors are unavailable for edge e. Thus, we can color edge e without any conflicts. This proves the statement for $n = k+1$ and hence by induction we get that the statement holds for all $n \geq 0$.

(c) We will use induction on the number of vertices n in the tree to prove the statement: For a tree with $n \geq 1$ vertices, if the maximum degree of any vertex is d, then the tree can be colored with d colors.
Base case (n=1). If there is only one vertex, then there are no edges to color, and thus can be colored with 0 colors.
Inductive hypothesis. Suppose the statement holds for $n = k \geq 1$.
Inductive Step. Remove any leaf v of your choice from the tree. We can then color the remaining tree with d colors by the inductive hypothesis. For any neighboring vertex u of vertex v,

the degree of u is at most $d-1$ since we removed the edge $\{u,v\}$ along with the vertex v. Thus its incident edges use at most $d-1$ colors and there is a color available for coloring the edge {u,v}. This completes the inductive step and by induction we have that the statement holds for all $n \geq 1$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 3B

1 Modular Inverses

Recall the definition of inverses from lecture: let $a, m \in \mathbb{Z}$ and $m > 0$; if $x \in \mathbb{Z}$ satisfies $ax \equiv 1 \pmod{m}$, then we say x is an **inverse of** a **modulo** m.

Now, we will investigate the existence and uniqueness of inverses.

(a) Is 3 an inverse of 5 modulo 10?

(b) Is 3 an inverse of 5 modulo 14?

(c) Is each $3 + 14n$ where $n \in \mathbb{Z}$ an inverse of 5 modulo 14?

(d) Does 4 have inverse modulo 8?

(e) Suppose $x, x' \in \mathbb{Z}$ are both inverses of a modulo m. Is it possible that $x \not\equiv x' \pmod{m}$?

Solution:

(a) No, because $3 \cdot 5 = 15 \equiv 5 \pmod{10}$.

(b) Yes, because $3 \cdot 5 = 15 \equiv 1 \pmod{14}$.

(c) Yes, because $(3 + 14n) \cdot 5 = 15 + 14 \cdot 5n \equiv 15 \equiv 1 \pmod{14}$.

(d) No. For contradiction, assume $x \in \mathbb{Z}$ is an inverse of 4 modulo 8. Then $4x \equiv 1 \pmod{8}$. Then $8 \mid 4x - 1$, which is impossible.

(e) No. We have $xa \equiv x'a \equiv 1 \pmod{m}$. So

$$xa - x'a = a(x - x') \equiv 0 \pmod{m}.$$

Multiply both sides by x, we get

$$xa(x - x') \equiv 0 \cdot x \pmod{m}$$

$$\implies x - x' \equiv 0 \pmod{m}.$$

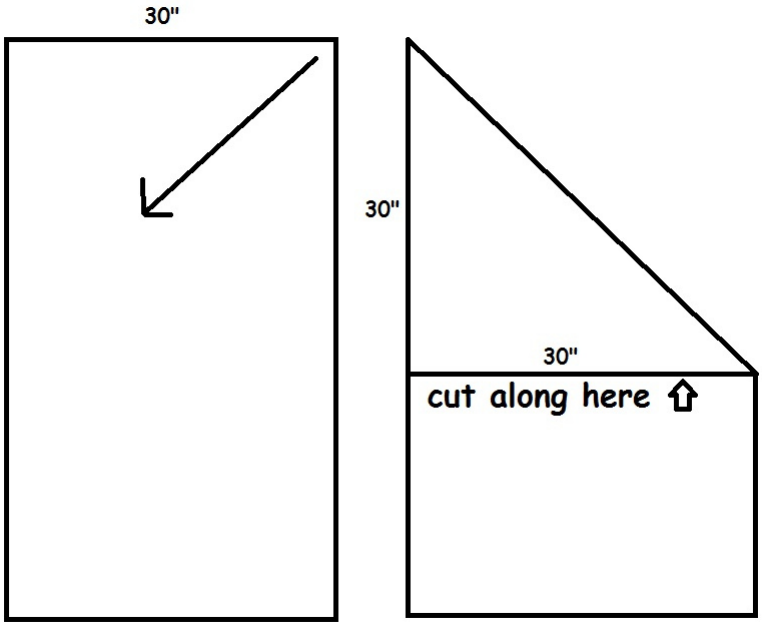
$$\implies x \equiv x' \pmod{m}$$

2 Paper GCD

Given a sheet of paper such as this one, and no rulers, describe a method to find the GCD of the width and the height of the paper. You can fold or tear the paper however you want, and ultimately you should produce a square piece whose side lengths are equal to the GCD.

Solution:

We can fold the smaller side diagonally onto the larger side, and tear the paper from where the fold lands.



If we started with height and width equal to a and b, this gives us a piece of paper with side lengths $a-b$ and b (assuming that $a > b$). Note that if $a-b > b$, the next time we end up with side lengths $a-2b$ and b. So after a few steps we must reach a mod b and b, at which we start subtracting from b.

Continuing this method is similar to the Euclidean algorithm and therefore results in reaching 0 at some point. Right before reaching 0, we must have a square piece of paper whose side lengths are the GCD.

3 Amaze Your Friends

It's been a long week, and you're finally in the Friday Zoom hangout that you've been looking forward to. You eschew conversations about Professor Rao's updated facial hair, that sourdough starter that's all the rage, or the new season of "Pose". Instead, you decide to invoke wonder (or possibly fear) in your friends by tricking them into thinking you can perform mental arithmetic with very large numbers.

So, what are the last digit of the following numbers?

(a) 11^{2017}

(b) 9^{10001}

(c) $3^{987654321}$

Solution:

(a) 11 is always 1 mod 10, so the answer to (a) is 1.

(b) 9 is its own inverse mod 10, therefore, if 9 is raised to an odd power, the number will be 9 mod 10. So the answer is 9.

Also notice that $9 \equiv -1 \pmod{10}$ so $9^{10001} \equiv (-1)^{10001} \equiv -1 \equiv 9 \pmod{10}$. In general, $m-1 \equiv -1 \pmod{m}$, so $m-1$ is always its own inverse. This is a useful trick so you can avoid computing the inverse of $m-1$ by hand. You can also check that $(m-1)^2 \equiv m^2 - 2m + 1 \equiv 1 \pmod{m}$, which is another proof that $m-1$ is its own inverse modulo m.

(c) $3^4 = 9^2 = 1$ mod 10. We see that the exponent $987654321 = 1$ mod 4 so the answer is 3.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 04A

1 Chinese Remainder Theorem Practice

In this question, you will solve for a natural number x such that,

$$\begin{aligned} x &\equiv 2 \pmod 3 \\ x &\equiv 3 \pmod 5 \\ x &\equiv 4 \pmod 7 \end{aligned} \tag{1}$$

(a) Suppose you find 3 natural numbers a, b, c that satisfy the following properties:

$$a \equiv 2 \pmod 3 \; ; \; a \equiv 0 \pmod 5 \; ; \; a \equiv 0 \pmod 7, \tag{2}$$
$$b \equiv 0 \pmod 3 \; ; \; b \equiv 3 \pmod 5 \; ; \; b \equiv 0 \pmod 7, \tag{3}$$
$$c \equiv 0 \pmod 3 \; ; \; c \equiv 0 \pmod 5 \; ; \; b \equiv 4 \pmod 7. \tag{4}$$

Show how you can use the knowledge of a, b and c to compute an x that satisfies (1).

In the following parts, you will compute natural numbers a, b and c that satisfy the above 3 conditions and use them to find an x that indeed satisfies (1).

(b) Find a natural number a that satisfies (2). In particular, an a such that $a \equiv 2 \pmod 3$ and is a multiple of 5 and 7. It may help to approach the following problem first:

(b.i) Find a^*, the multiplicative inverse of 5×7 modulo 3. What do you see when you compute $(5 \times 7) \times a^*$ modulo 3, 5 and 7? What can you then say about $(5 \times 7) \times (2 \times a^*)$?

(c) Find a natural number b that satisfies (3). In other words: $b \equiv 3 \pmod 5$ and is a multiple of 3 and 7.

(d) Find a natural number c that satisfies (4). That is, c is a multiple of 3 and 5 and $\equiv 4 \pmod 7$.

(e) Putting together your answers for Part (a), (b), (c) and (d), report an x that indeed satisfies (1).

Solution:

(a) Observe that $a+b+c \equiv 2+0+0 \pmod 3$, $a+b+c \equiv 0+3+0 \pmod 5$ and $a+b+c \equiv 0+0+4 \pmod 7$. Therefore $x = a+b+c$ indeed satisfies the conditions in (1).

(b) This question asks to find a number $0 \leq a < 3 \times 5 \times 7$ that is divisible by 5 and 7 and returns 2 when divided by 3. Let's first look at Part (b.i):

(b.i) Observe that $(5 \times 7) \equiv 35 \equiv 2 \pmod 3$. Multiplying both sides by 2, this means that $2 \times (5 \times 7) \equiv 4 \pmod 3 \equiv 1 \pmod 3$. So, the multiplicative inverse of 5×7, a^* is exactly 2. To verify this: observe that $(5 \times 7) \times 2 = 70 = 3 \times 23 + 1$. Therefore $(5 \times 7) \times 2 \equiv 1 \pmod 3$.

Consider $5 \times 7 \times a^*$. Since it is a multiple of 5 and 7, it is equal to 0 modulo either of these numbers. On the other hand, $5 \times 7 \times a^* \equiv 1 \pmod 3$, since a^* is precisely defined to be the multiplicative inverse of 5×7 modulo 3.

Consider $5 \times 7 \times (2 \times a^*) = 140$. It is a multiple of, and is therefore 0 modulo both 5 and 7. On the other hand, $5 \times 7 \times (2 \times a^*) \equiv 1 \times 2 \pmod 3$, for the same reason that a^* is defined to be the multiplicative inverse of 5×7 modulo 3.

Indeed observe that $5 \times 7 \times (2 \times a^*) = 140$ precisely satisfies the criteria required in Part (b). It is equivalent to 0 modulo 5 and 7 and $\equiv 2 \pmod 3$.

(c) Let's try to use a similar approach as Part (b). In particular, first observe that $3 \times 7 \equiv 21 \equiv 1 \pmod 5$. Therefore, b^*, the multiplicative inverse of 3×7 modulo 5 is in fact 1! So, let us consider $3 \times 7 \times (3 \times b^*) = 63$: this is a multiple of 3 and 7 and is therefore 0 modulo both these numbers. On the other hand, $3 \times 7 \times (3 \times b^*) \equiv 3 \pmod 5$ for the reason that b^* is the multiplicative inverse of 3×7 modulo 5.

(d) Yet again the approach of Part (b) proves to be useful! Observe that $3 \times 5 \equiv 15 \equiv 1 \pmod 7$. Therefore, c^*, the multiplicative inverse of 3×5 modulo 7 turns out to be 1. So, let us consider $3 \times 5 \times (4 \times c^*) = 60$: this is a multiple of 3 and 5. is therefore 0 modulo both these numbers. On the other hand, $3 \times 5 \times (4 \times c^*) \equiv 4 \pmod 7$ for the reason that c^* is the multiplicative inverse of 3×5 modulo 7.

(e) From Parts (b), (c) and (d) we find a choice of a, b, c (respectively $= 140, 63, 60$) which satisfues (2), (3) and (4). Together with Part (a) of the question, this implies that $x = a + b + c = 263$ satisfies the required criterion in (1).

To verify this: observe that,

$$263 = 87 \times 3 + 2,$$
$$263 = 52 \times 5 + 3,$$
$$263 = 37 \times 7 + 4.$$

2 CRT Decomposition

In this problem we will find $3^{302} \mod 385$.

(a) Write 385 as a product of prime numbers in the form $385 = p_1 \times p_2 \times p_3$.

(b) Use Fermat's Little Theorem to find $3^{302} \mod p_1$, $3^{302} \mod p_2$, and $3^{302} \mod p_3$.

(c) Let $x = 3^{302}$. Use part (b) to express the problem as a system of congruences (modular equations $\mod 385$). Solve the system using the Chinese Remainder Theorem. What is $3^{302} \mod 385$?

Solution:

(a) $385 = 11 \times 7 \times 5$.

(b) Since $3^4 \equiv 1 \pmod 5$, $3^{302} \equiv 3^{4(75)} \cdot 3^2 \equiv 4 \pmod 5$.
Since $3^6 \equiv 1 \pmod 7$, $3^{302} \equiv 3^{6(50)} \cdot 3^2 \equiv 2 \pmod 7$.
Since $3^{10} \equiv 1 \pmod{11}$, $3^{302} \equiv 3^{10(30)} \cdot 3^2 \equiv 9 \pmod{11}$.

(c) $x \equiv 4 \pmod 5$, $x \equiv 2 \pmod 7$, $x \equiv 9 \pmod{11}$.
The answer we get using CRT is $x \equiv 9 \pmod{385}$. So $3^{302} \equiv 9 \pmod{385}$.

3 Baby Fermat

Assume that a does have a multiplicative inverse mod m. Let us prove that its multiplicative inverse can be written as $a^k \pmod m$ for some $k \geq 0$.

(a) Consider the sequence $a, a^2, a^3, \ldots \pmod m$. Prove that this sequence has repetitions. (**Hint:** Consider the Pigeonhole Principle.)

(b) Assuming that $a^i \equiv a^j \pmod m$, where $i > j$, what can you say about $a^{i-j} \pmod m$?

(c) Prove that the multiplicative inverse can be written as $a^k \pmod m$. What is k in terms of i and j?

Solution:

(a) There are only m possible values mod m, and so after the m-th term we should see repetitions.
The Pigeonhole principle applies here - we have m boxes that represent the different unique values that a^k can take on $\pmod m$. Then, we can view $a, a^2, a^3, \cdots$ as the objects to put in the m boxes. As soon as we have more than m objects (in other words, we reach a^{m+1} in our sequence), the Pigeonhole Principle implies that there will be a collision, or that at least two numbers in our sequence take on the same value $\pmod m$.

(b) We will temporarily use the notation a^* for the multiplicative inverse of a to avoid confusion.

If we multiply both sides by $(a^*)^j$ in the third line below, we get

$$\begin{aligned}
a^i &\equiv a^j && \pmod m,\\
a^{i-j}\underbrace{a\cdots a}_{j \text{ times}} &\equiv \underbrace{a\cdots a}_{j \text{ times}} && \pmod m,\\
a^{i-j}\underbrace{a\cdots a}_{j \text{ times}}\cdot\underbrace{a^*\cdots a^*}_{j \text{ times}} &\equiv \underbrace{a\cdots a}_{j \text{ times}}\cdot\underbrace{a^*\cdots a^*}_{j \text{ times}} && \pmod m,\\
a^{i-j} &\equiv 1 && \pmod m.
\end{aligned}$$

(c) We can rewrite $a^{i-j} \equiv 1 \pmod m$ as $a^{i-j-1}a \equiv 1 \pmod m$. Therefore a^{i-j-1} is the multiplicative inverse of $a \pmod m$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 04B

1 RSA Practice

Consider the following RSA schemes and solve for asked variables.

(a) Assume for an RSA scheme we pick 2 primes $p = 5$ and $q = 11$ with encryption key $e = 9$, what is the decryption key d? Calculate the exact value.

(b) If the receiver gets 4, what was the original message?

(c) Encode your answer from part (b) to check its correctness.

Solution:

(a) The private key d is defined as the inverse of $e \pmod{(p-1)(q-1)}$. Thus we need to compute $9^{-1} \bmod (5-1)(11-1) = 9^{-1} \bmod 40$. Find inverse of e mod $(5-1)(11-1) = 40$. Compute egcd$(40,9)$:

$$\begin{aligned} \text{egcd}(40,9) &= \text{egcd}(9,4) && [4 = 40 \bmod 9 = 40 - 4(9)] \\ &= \text{egcd}(4,1) && [1 = 9 \bmod 4 = 9 - 2(4)]. \\ 1 &= 9 - 2(4). \\ 1 &= 9 - 2(40 - 4(9)) \\ &= 9 - 2(40) + 8(9) = 9(9) - 2(40). \end{aligned}$$

We get $-2(40) + 9(9) = 1$. So the inverse of 9 is 9. So $d = 9$.

(b) 4 is the encoded message. We can decode this with $D(m) \equiv m^d \equiv 4^9 \equiv 14 \pmod{55}$. $4^9 \equiv 14 \pmod{55}$. Thus the original message was 14.

(c) The answer from the second part was 14. To encode the number x we must compute $x^e \bmod N$. Thus, $14^9 \equiv 14 \cdot (14^2)^4 \equiv 14 \cdot (31^2)^2 \equiv 14 \cdot (26^2) \equiv 14 \cdot 16 \equiv 4 \pmod{55}$. This verifies the second part since the encoded message was suppose to be 4.

2 RSA Practice

Bob would like to receive encrypted messages from Alice via RSA.

(a) Bob chooses $p = 7$ and $q = 11$. His public key is (N, e). What is N?

(b) What number is e relatively prime to?

(c) e need not be prime itself, but what is the smallest prime number e can be? Use this value for e in all subsequent computations.

(d) What is $\gcd(e,(p-1)(q-1))$?

(e) What is the decryption exponent d?

(f) Now imagine that Alice wants to send Bob the message 30. She applies her encryption function E to 30. What is her encrypted message?

(g) Bob receives the encrypted message, and applies his decryption function D to it. What is D applied to the received message?

Solution:

(a) $N = pq = 77$.

(b) e must be relatively prime to $(p-1)(q-1) = 60$.

(c) We cannot take $e = 2, 3, 5$, so we take $e = 7$.

(d) By design, $\gcd(e,(p-1)(q-1)) = 1$ always.

(e) The decryption exponent is $d = e^{-1} \pmod{60} = 43$, which could be found through Euclid's extended GCD algorithm.

(f) The encrypted message is $E(30) = 30^7 \equiv 2 \pmod{77}$. We can obtain this answer via repeated squaring.

$$\begin{aligned} 30^7 &\equiv 30 \cdot 30^6 \equiv 30 \cdot (30^2 \bmod 77)^3 \equiv 30 \cdot 53^3 \equiv (30 \cdot 53 \bmod 77) \cdot (53^2 \bmod 77) \equiv 50 \cdot 37 \\ &\equiv 2 \pmod{77}. \end{aligned}$$

(g) We have $D(2) = 2^{43} \equiv 30 \pmod{77}$. Again, we can use repeated squaring.

$$\begin{aligned} 2^{43} &\equiv 2 \cdot 2^{42} \equiv 2 \cdot (2^2 \bmod 77)^{21} \equiv 2 \cdot 4^{21} \equiv (2 \cdot 4 \bmod 77) \cdot 4^{20} \equiv 8 \cdot (4^2 \bmod 77)^{10} \\ &\equiv 8 \cdot 16^{10} \equiv 8 \cdot (16^2 \bmod 77)^5 \equiv 8 \cdot 25^5 \equiv (8 \cdot 25 \bmod 77) \cdot 25^4 \equiv 46 \cdot (25^2 \bmod 77)^2 \\ &\equiv 46 \cdot (9^2 \bmod 77) \equiv 46 \cdot 4 \equiv 30 \pmod{77}. \end{aligned}$$

3 RSA Lite

Woody misunderstood how to use RSA. So he selected prime $P = 101$ and encryption exponent $e = 67$, and encrypted his message m to get $35 = m^e \bmod P$. Unfortunately he forgot his original message m and only stored the encrypted value 35. But Carla thinks she can figure out how to

recover m from $35 = m^e \mod P$, with knowledge only of P and e. Is she right? Can you help her figure out the message m? Show all your work.

Solution:

Recall that the security of RSA depended upon the supposed hardness of factoring $N = P \times Q$. However, since $N = P$ in this problem, we can consider it to have been already factored! Indeed, recall that the private key d in RSA is defined to be the multiplicative inverse of e modulo $(P-1)(Q-1)$, because we can then use the following relation to decrypt the message:

$$m^{k(P-1)(Q-1)+1} \equiv m \pmod{N}$$

Note that in our case where $N = P$, an analogous relation immediately holds by Fermat's Little Theorem:

$$m^{k(P-1)+1} \equiv m \pmod{P}$$

Therefore, if we can find d which is the multiplicative inverse of e modulo $P-1$, we can decrypt the message by simply computing $m^{ed} \pmod{P} = 35^d \pmod{P}$. It is easy to see by inspection that $67 \times 3 = 201 \equiv 1 \pmod{100}$, so the desired multiplicative inverse $d = 3$, which means that $m = 51 \pmod{101}$.

(Otherwise, one can find the multiplicative inverse by applying Extended Euclid's algorithm to $e = 67$ and $P - 1 = 100$:

$$\begin{aligned}
(c,a,b) &= \text{extended-gcd}(100,67) = (c, b_1, a_1 - \lfloor 100/67 \rfloor b_1) \quad \text{where} \\
(c,a_1,b_1) &= \text{extended-gcd}(67,33) = (c, b_2, a_2 - \lfloor 67/33 \rfloor b_2) \quad \text{where} \\
(c,a_2,b_2) &= \text{extended-gcd}(33,1) = (c, b_3, a_3 - \lfloor 33/1 \rfloor b_3) \quad \text{where} \\
(c,a_3,b_3) &= \text{extended-gcd}(1,0) = (1,1,0)
\end{aligned}$$

Therefore, $(c,a_2,b_2) = (1,0,1)$, $(c,a_1,b_1) = (1,1,-2)$, and $(c,a,b) = (1,-2,3)$ respectively. We can verify that $1 = c = ax + by = -2 \times 100 + 3 \times 67$. Hence, the multiplicative inverse of 67 modulo 100 is 3.)

4 RSA with Three Primes

Show how you can modify the RSA encryption method to work with three primes instead of two primes (i.e. $N = pqr$ where p, q, r are all prime), and prove the scheme you come up with works in the sense that $D(E(x)) \equiv x \pmod{N}$.

Solution:

$N = pqr$ where p, q, r are all prime. Then, let e be co-prime with $(p-1)(q-1)(r-1)$. Give the public key: (N, e) and calculate $d = e^{-1} \mod (p-1)(q-1)(r-1)$. People who wish to send me a secret, x, send $y = x^e \mod N$. I decrypt an incoming message, y, by calculating $y^d \mod N$.

Does this work? We need to prove that $x^{ed} - x \equiv 0 \pmod{N}$ and thus $x^{ed} \equiv x \pmod{N}$. To prove that $x^{ed} - x \equiv 0 \pmod{N}$, we factor out the x to get $x \cdot (x^{ed-1} - 1) = x \cdot (x^{k(p-1)(q-1)(r-1)+1-1} - 1)$

because $ed \equiv 1 \pmod{(p-1)(q-1)(r-1)}$. As a reminder, we are considering the number: $x \cdot (x^{k(p-1)(q-1)(r-1)} - 1)$.

We now argue that this number must be divisible by p, q, and r. Thus it is divisible by N and $x^{ed} - x \equiv 0 \pmod{N}$.
To prove that it is divisible by p:

- If x is divisible by p, then the entire thing is divisible by p.
- If x is not divisible by p, then that means we can use FLT on the inside to show that $(x^{p-1})^{k(q-1)(r-1)} - 1 \equiv 1 - 1 \equiv 0 \pmod{p}$. Thus it is divisible by p.

The same reasoning shows that it is divisible by q and r.

One can also use a CRT based argument to argue the correctness of 3 prime RSA. Indeed, as discussed in the previous paragraphs, we need to show that $x^{ed} \equiv x \pmod{N}$, where recall that $N = pqr$. In order to do this, observe that it suffices to prove the following three equivalences:

$$x^{ed} \equiv x \pmod{p}, \tag{1}$$
$$x^{ed} \equiv x \pmod{q}, \tag{2}$$
$$x^{ed} \equiv x \pmod{r}. \tag{3}$$

Why does it suffice? If these 3 statements are indeed true, the uniqueness property established in the CRT implies that $x^{ed} \equiv x \mod N$. Note that p, q and r are relatively prime so we are allowed to apply the Chinese Remainder Theorem here.

Recall that $e > 1$ is any natural number that is relatively prime to $p-1$, $q-1$ and $r-1$. And d is the multiplicative inverse of e modulo $(p-1)(q-1)(r-1)$. In particular, this means that $ed = k(p-1)(q-1)(r-1)+1$ for some natural number k. Let us try to use this to verify (1):

$$\begin{aligned} x^{ed} &= x^{k(p-1)(q-1)(r-1)+1} \\ &= x \cdot \left(x^{k(q-1)(r-1)}\right)^{p-1} \\ &\equiv x \pmod{p} \end{aligned}$$

where the last step follows by using Fermat's Little Theorem to claim that for any $a \in \mathbb{N}$, $a^{p-1} \equiv 1 \pmod{p}$. In particular, we choose $a = x^{k(q-1)(r-1)}$ and apply FLT. Note that the original FLT holds with $a = 1, 2, \cdots, p-1$, but we leave it as an exercise to prove that it indeed applies for any natural number $a \in \mathbb{N}$. Thus, we have shown that $x^{ed} \equiv x \pmod{p}$, and a matching argument shows that $x^{ed} \equiv x \pmod{q}$ and $x^{ed} \equiv x \pmod{r}$. This proves equations (1), (2) and (3) and hence shows that $x^{ed} \equiv x \pmod{N}$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 Dis 05A

1 Polynomial Practice

(a) If f and g are non-zero real polynomials, how many roots do the following polynomials have at least? How many can they have at most? (Your answer may depend on the degrees of f and g.)

(i) $f+g$

(ii) $f \cdot g$

(iii) f/g, assuming that f/g is a polynomial

(b) Now let f and g be polynomials over GF(p).

(i) We say a polynomial $f = 0$ if $\forall x, f(x) = 0$. If $f \cdot g = 0$, is it true that either $f = 0$ or $g = 0$?

(ii) How many f of degree *exactly* $d < p$ are there such that $f(0) = a$ for some fixed $a \in \{0, 1, \ldots, p-1\}$?

(c) Find a polynomial f over GF(5) that satisfies $f(0) = 1, f(2) = 2, f(4) = 0$. How many such polynomials are there?

Solution:

(a) (i) It could be that $f+g$ has no roots at all (example: $f(x) = 2x^2 - 1$ and $g(x) = -x^2 + 2$), so the minimum number is 0. However, if the highest degree of $f+g$ is odd, then it has to cross the x-axis at least once, meaning that the minimum number of roots for odd degree polynomials is 1 (we did not look for this case when grading). On the other hand, $f+g$ is a polynomial of degree at most $m = \max(\deg f, \deg g)$, so it can have at most m roots. The one exception to this expression is if $f = -g$. In that case, $f+g = 0$, so the polynomial has an infinite number of roots!

(ii) A product is zero if and only if one of its factors vanishes. So if $f(x) \cdot g(x) = 0$ for some x, then either x is a root of f or it is a root of g, which gives a maximum of $\deg f + \deg g$ possibilities. Again, there may not be any roots if neither f nor g have any roots (example: $f(x) = g(x) = x^2 + 1$).

(iii) If f/g is a polynomial, then it must be of degree $d = \deg f - \deg g$ and so there are at most d roots. Once more, it may not have any roots, e.g. if $f(x) = g(x)(x^2+1)$, $f/g = x^2 + 1$ has no root.

(b) (i) **Example 1:** $x^{p-1}-1$ and x are both non-zero polynomials on $GF(p)$ for any p. x has a root at 0, and by Little Fermat, $x^{p-1}-1$ has a root at all non-zero points in $GF(p)$. So, their product $x^p - x$ must have a zero on all points in $GF(p)$.
Example 2: To satisfy $f \cdot g = 0$, all we need is $(\forall x \in S, f(x) = 0 \vee g(x) = 0)$ where $S = \{0, \ldots, p-1\}$. We may see that this is not equivalent to $(\forall x \in S, f(x) = 0)) \vee (\forall x \in S, g(x) = 0)$.
To construct a concrete example, let $p = 2$ and we enforce $f(0) = 1, f(1) = 0$ (e.g. $f(x) = 1 - x$), and $g(0) = 0, g(1) = 1$ (e.g. $g(x) = x$). Then $f \cdot g = 0$ but neither f nor g is the zero polynomial.

(ii) We know that in general each of the $d+1$ coefficients of $f(x) = \sum_{k=0}^{d} c_k x^k$ can take any of p values. However, the conditions $f(0)$ and $\deg f = d$ impose constraints on the constant coefficient $f(0) = c_0 = a$ and the top coefficient $x_d \neq 0$. Hence we are left with $(p-1) \cdot p^{d-1}$ possibilities.

(c) We know by part (b) that any polynomial over GF(5) can be of degree at most 4. A polynomial of degree ≤ 4 is determined by 5 points (x_i, y_i). We have assigned three, which leaves $5^2 = 25$ possibilities. To find a specific polynomial, we use Lagrange interpolation:

$$\Delta_0(x) = 2(x-2)(x-4) \qquad \Delta_2(x) = x(x-4) \qquad \Delta_4(x) = 2x(x-2),$$

and so $f(x) = \Delta_0(x) + 2\Delta_2(x) = 4x^2 + 1$.

2 Rational Root Theorem

The rational root theorem states that for a polynomial

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

$a_0, \cdots, a_n \in \mathbb{Z}$, if $a_0, a_n \neq 0$, then for each rational solution $\frac{p}{q}$ such that $\gcd(p,q) = 1$, $p | a_0$ and $q | a_n$. Prove the rational root theorem.

Solution: If $\frac{p}{q}$ is a root of the polynomial P, we can write

$$P\left(\frac{p}{q}\right) = a_n \left(\frac{p}{q}\right)^n + \cdots + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Multiplying both sides by q^n we get

$$p(a_n p^{n-1} + a_{n-1} q p^{n-1} + \cdots + a_1 q^{n-1}) = -a_0 q^n$$

From this we can see that p divides $a_0 q^n$; however, recall that p and q are coprime, so p must divide a_0, as desired.

If instead we chose to factor out q, we have

$$q(a_{n-1} p^{n-1} + \cdots + a_0 q^{n-1}) = -a_n p^n$$

and for the same reasons we can say that q divides a_n.

3 Secret Sharing Practice

Consider the following secret sharing schemes and solve for asked variables.

(a) Suppose there is a bag of candy locked with a passcode between 0 and an integer n. Create a scheme for 5 trick-or-treaters such that they can only open the bag of candy if 3 of them agree to open it.

(b) Create a scheme for the following situation: There are 4 cats and 3 dogs in the neighborhood, and you want them to only be able to get the treats if the majority of the animals of each type are hungry. The treats are locked by a passcode between 0 and an integer n.

Solution:

(a) Solutions vary. The polynomial should be degree 2 and each trick-or-treater should be given the polynomial evaluated at one point.

(b) The guiding principle in this solution is that a polynomial of degree d, is uniquely determined by $d+1$ points. Let there be three polynomials, one for cats c, one for dogs d, and one joint one j that has the secret that actually unlocks the treats. c will be degree 2 since you need 3 cats to agree to get the 3 points to uniquely determine it. and d with be degree 1 since you need 2 dogs to agree to get the 2 points to uniquely determine it. The j will be degree 1 and $c(0)$ will be $j(1)$, and the $d(0)$ will be $j(2)$. This way you need both the point from the dogs and the point from the cats to uniquely determine j and otherwise you will be unable to determine the $j(0)$. This is also why we make $j(0)$ our secret.

4 Old Secrets, New Secrets

In order to share a secret number s, Alice distributed the values $(1,p(1)),(2,p(2)),\ldots,(n+1,p(n+1))$ of a degree n polynomial p with her friends $\text{Bob}_1,\ldots,\text{Bob}_{n+1}$. As usual, she chose p such that $p(0)=s$. Bob_1 through Bob_{n+1} now gather to jointly discover the secret. Suppose that for some reason Bob_1 already knows s, and wants to play a joke on $\text{Bob}_2,\ldots,\text{Bob}_{n+1}$, making them believe that the secret is in fact some fixed $s' \neq s$. How could he achieve this? In other words, what value should he report in order to make the others believe that the secret is s'?

Solution:

We know that in order to discover s, the Bobs would compute

$$s = y_1\Delta_1(0) + \sum_{k=2}^{n+1} y_k\Delta_k(0), \tag{1}$$

where $y_i = p(i)$. Bob_1 now wants to change his value y_1 to some y_1', so that

$$s' = y_1'\Delta_1(0) + \sum_{k=2}^{n+1} y_k\Delta_k(0). \tag{2}$$

Subtracting Equation 1 from 2 and solving for y_1', we see that

$$y_1' = (\Delta_1(0))^{-1}\left(s' - s\right) + y_1,$$

where $(\Delta_1(0))^{-1}$ exists, because $\deg \Delta_1(x) = n$ with its n roots at $2, \ldots, n+1$ (so $\Delta_1(0) \neq 0$).

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 06A

1 Zerg Player

A Zerg player wants to produce an army to fight against Protoss in early game, and he wants to have a small army which consumes exactly 10 supply. And he has the following choices:

- Zerglings: consumes 1 supply
- Hydralisk: consumes 2 supply
- Roach: consumes 2 supply

How many different compositions can the player's army have? Note that Zerglings are indistinguishable, as are Hydralisks and Roachs.

Solution: Let there are i 2-supply units have been made. For the rest of supply, we can fill it with zerglings.

And if there are i 2-supply unites, there are $i+1$ different compositions: 0 Hydra i Roach $10-2i$ zerglings, 1 Hydra $i-1$ Roach $10-2i$ zerglings, ..., i Hydra 0 Roach $10-2i$ zerglings.

Then we have $\sum_{i=0}^{5}(i+1) = 1+2+3+4+5+6 = 21$.

2 Strings

What is the number of strings you can construct given:

(a) n ones, and m zeroes?

(b) n_1 A's, n_2 B's and n_3 C's?

(c) $n_1, n_2, \ldots, n_k$ respectively of k different letters?

Solution:

(a) $\binom{n+m}{n}$

(b) $(n_1+n_2+n_3)!/(n_1!\cdot n_2!\cdot n_3!)$

(c) $(n_1+n_2+\cdots+n_k)!/(n_1!\cdot n_2!\cdots n_k!)$.

3 Counting Game

RPG games are all about explore different mazes. Here is a weird maze: there are 2^n rooms, where each room is the vertex on a the n-dimensional hypercube, labeled by a n bit binary string.

For each room, there are n different doors, each door corresponding to an edge on the hypercube. If you are at room i, and choose door j, then you will go to room $i \oplus 2^j$ (flips the $j+1$-th bit in number i).

(a) How many different shortest path are there from room 0 to room $2^n - 1$?

(b) How many different paths of $n+2$ steps are there to go from room 0 to room $2^n - 1$?

(c) If $n = 8$, how many different shortest pathes are there from room 0 to room 63 that pass through 3 and 19?

Solution:

(a) $n!$, the shortest path is n, and for the i-th step, there are only $n-i$ doors flips a zero to one.

(b) The player made one mistake during his trip, so suppose he made the mistake at step i, $i > 0$, so there are i different ways to make the mistake. Then he will start from a room with $n-i+1$ zeros. So the total number is $\sum_{i=1}^{n}(\binom{n}{i} * i! * i * (n-i+1)!)$.

Optional for further steps:

$\sum_{i=1}^{n}(\binom{n}{i} * i! * i * (n-i+1)!) = \sum_{i=1}^{n} \frac{n!*i!*i*(n-i+1)!}{(n-i)!i!)} = \sum_{i=1}^{n} n! * i * (n-i+1) = n! \sum_{i=1}^{n} i(n-i+1) = n! * (\sum_{i=1}^{n}(in - i^2 + i))$ where $\sum_{i=1}^{n}(in - i^2 + i) = n * \sum_{i=1}^{n} i - \sum_{i=1}^{n} i^2 + \sum_{i=1}^{n} i = \frac{n(n+1)(n+2)}{6}$

(c) From 0 to 3, 2 different pathes. From 3 to 19: notice $3 \oplus 19 = 16$ so there is only one way. From 19 to 63, there are 3 zeros in $63 \oplus 19$ so total 3! different pathes. In total $2 * 3!$ different pathes.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 6B

1 Count it

Let's get some practice with counting!

(a) How many sequences of 15 coin-flips are there that contain exactly 4 heads?

(b) An anagram of HALLOWEEN is any re-ordering of the letters of HALLOWEEN, i.e., any string made up of the letters H, A, L, L, O, W, E, E, N in any order. The anagram does not have to be an English word.
How many different anagrams of HALLOWEEN are there?

(c) How many solutions does $y_0 + y_1 + \cdots + y_k = n$ have, if each y must be a non-negative integer?

(d) How many solutions does $y_0 + y_1 = n$ have, if each y must be a positive integer?

(e) How many solutions does $y_0 + y_1 + \cdots + y_k = n$ have, if each y must be a positive integer?

Solution:

(a) This is just the number of ways to choose 4 positions out of 15 positions to place the heads, and so is $\binom{15}{4}$.

(b) In this 9 letter word, the letters L and E are each repeated 2 times while the other letters appear once. Hence, the number 9! overcounts the number of different anagrams by a factor of $2! \times 2!$ (one factor of 2! for the number of ways of permuting the 2 L's among themselves and another factor of 2! for the number of ways of permuting the 2 E's among themselves). Hence, there are $9!/(2!)^2$ different anagrams.

(c) $\binom{n+k}{k}$. We can imagine this as a sequence of n ones and k plus signs: y_0 is the number of ones before the first plus, y_1 is the number of ones between the first and second plus, etc. We can now count the number of sequences using the "balls and bins" method (also known as "stars and bars").

(d) $n-1$. We can just enumerate the solutions here. y_0 can take values $1, 2, \ldots, n-1$ and this uniquely fixes the value of y_1. So, we have $n-1$ ways to do this. But, this is just an example of the more general question below.

(e) $\binom{(n-(k+1))+k}{k} = \binom{n-1}{k}$. By subtracting 1 from all $k+1$ variables, and $k+1$ from the total required, we reduce it to problem with the same form as the previous problem. Once we have

a solution to that we reverse the process, and adding 1 to all the non-negative variables gives us positive variables.

2 Inclusion and exclusion

What is total number of positive numbers that smaller than 100 and coprime to 100?

Solution: It's enough to count the inverse: what is the total number of positive ingeters that smaller than 100 and not coprime to 100?

Not coprime to 100 means that the number either is a multiple of 2 or a multiple of 5. Then we have

49 numbers are multiple of 2. 19 numbers are multiple of 5. 9 numbers that are multiple of both 2 and 5.

So the total number is $49+19-9=59$, and there are 99 positive integers smaller than 100.

So in total, there are $99-59=40$ different number of positive numbers (smaller than 100) that are coprime to 100.

3 Identities

(a) $\sum_{i=0}^{n}(-1)^i\binom{n}{i}=0$

(b) $\sum_{i=0}^{n}\binom{r+i}{i}=\binom{r+n+1}{n}$

(c) $\sum_{i=0}^{n}\binom{r}{i}\binom{s}{n-i}=\binom{r+s}{n}$ (Note: Assuming $r>n, s>n$)

Solution:

(a) $0=(1-1)^n=\sum_{i=0}^{n}(-1)^i 1^{n-1}\binom{n}{i}$

(b) RHS $\binom{r+n+1}{n}$ can be viewed as counting the number of subsets of $\{1,2,...,n+r+1\}$ of size n. $\sum_{i=1}^{n}\binom{r+i}{i}$ can be viewed as counting the same thing but in a different way. It first specifies the smallest element that is 'NOT' in the selected subset. For example if 1 is NOT in the subset then there are $\binom{n+r+1-1}{n}$ ways of different subsets. If 2 is the smallest that not in the subset then 1 is in the subset and there are $\binom{n+r+1-2}{n-1}$ different ways remaining, etc. down to if $n+1$ is the smallest number that not in the set, then $1,2,3,...,n$ are in the subset then we only have $\binom{n+r+1-(n+1)}{0}$ different ways. So RHS and LHS counts the same thing.

(c) RHS counts the total number of different subsets from $\{1,2,3,4,...,r,r+1,...,r+s\}$ that has size n, LHS counts the same thing by specify how many elements is selected from $\{1,2,3,4,...,r\}$ and how many of them are selected from $\{r+1,r+2,...,r+s\}$. If i of them are selected from the first set, then $n-i$ of them must be selected from the second set, the total number is $\binom{r}{i}\binom{s}{n-i}$. And we iterate through all possible i.

4 Largest binom

For which value(s) of k is $\binom{n}{k}$ maximum? Prove your answer.

Solution: When n is odd, $\binom{n}{\frac{n+1}{2}}$ and $\binom{n}{\frac{n+1}{2}-1}$ are maximum. When n is even, $\binom{n}{n/2}$ is maximum.

To prove this, we need the following equality:

$$\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$$

Proof: $\frac{n-k+1}{k}\binom{n}{k-1} = \frac{n!}{(k-1)!(n-k+1)!}\frac{n-k+1}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$

We noticed that $\frac{n-k+1}{k} \geq 1$ when $k \leq \frac{n+1}{2}$ for n is odd case, and $\frac{n}{2}$ for n is even case.

So $\binom{n}{k} \geq \binom{n}{k-1}$ when $k \leq \frac{n+1}{2}$ for n is odd case, and $\frac{n}{2}$ for n is even case.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 8A

1 Graph Isomorphic

In graph theory, an isomorphism of graphs G and H is a bijection between the vertex sets of G and H

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

such that any two vertices u, v of G are adjacent in G if and only if $f(u), f(v)$ are adjacent in H.

Prove the following:

1. The degrees of corresponding nodes $u, f(u)$ are the same.
2. There is a bijection between edges.
3. If G is connected, then H is also connected.

Solution:

(a) by definition, (u,v) in G **if and only if** $(f(u), f(v))$ in H. So they have the same degree.

(b) There is a bijection: (u,v) maps to $(f(u), f(v))$.

(c) If G is connected and H is not connected. Then there exists a pair of points u, v are connected in G by some path, but $f(u), f(v)$ are not connected in H by any path. Let the path be $u -> x_1 -> x_2, ..., -> x_k -> v$ where $k \geq 0$ ($k = 0$ means there is an edge between u, v). Since (u, x_1) is connected by some edge in G, then $(f(u), f(x_1))$ is connected by some edge in H. Similar argument holds for any adjacent pairs $(x_i, x_{i+1}), (f(x_i), f(x_{i+1}))$ or $(x_k, v), (f(x_k), f(v))$. So there is a path between $f(u)$ and $f(v)$ in H. Contradiction!

2 Countability Practice

(a) Do $(0,1)$ and $\mathbb{R}_+ = (0,\infty)$ have the same cardinality? If so, either give an explicit bijection (and prove that it is a bijection) or provide an injection from $(0,1)$ to $(0,\infty)$ and an injection from $(0,\infty)$ to $(0,1)$ (so that by Cantor-Bernstein theorem the two sets will have the same cardinality). If not, then prove that they have different cardinalities.

(b) Is the set of strings over the English alphabet countable? (Note that the strings may be arbitrarily long, but each string has finite length. Also the strings need not be real English words.) If so, then provide a method for enumerating the strings. If not, then use a diagonalization argument to show that the set is uncountable.

(c) Consider the previous part, except now the strings are drawn from a countably infinite alphabet $\mathscr{A}$. Does your answer from before change? Make sure to justify your answer.

Solution:

(a) Yes, they have the same cardinality.
Explicit bijection: Consider the bijection $f : (0,1) \to (0,\infty)$ given by

$$f(x) = \frac{1}{x} - 1.$$

We show that f is a bijection by proving separately that it is one-to-one and onto. The function f is one-to-one: suppose that $f(x) = f(y)$. Then,

$$\begin{aligned} \frac{1}{x} - 1 &= \frac{1}{y} - 1, \\ \frac{1}{x} &= \frac{1}{y}, \\ x &= y. \end{aligned}$$

Hence, f is one-to-one.

The function f is onto: take any $y \in (0,\infty)$. Let $x = 1/(1+y)$. Note that $x \in (0,1)$. Then,

$$f(x) = \frac{1}{1/(1+y)} - 1 = 1 + y - 1 = y,$$

so f maps x to y. Hence, f is onto.

We have exhibited a bijection from $(0,1)$ to $(0,\infty)$, so they have the same cardinality. (In fact, they are both uncountable.)

Indirect bijection: The injection from $(0,1)$ to $(0,\infty)$ is trivial; consider the function $f : (0,1) \to (0,\infty)$ given by

$$f(x) = x.$$

It is easy to see that f is injective.

For the other way, consider the function $g : (0,\infty) \to (0,1)$ given by

$$g(x) = \frac{1}{x+1}.$$

To see that g is injective, suppose $g(x) = g(y)$. Then

$$\frac{1}{x+1} = \frac{1}{y+1} \implies x = y.$$

Hence g is injective. Thus we have an injective function from $(0,1)$ to $(0,\infty)$ and an injective function from $(0,\infty)$ to $(0,1)$. By Cantor-Bernstein theorem there exists a bijection from $(0,1)$ to $(0,\infty)$ and hence they have the same cardianlity.

(b) Countable. The English language has a finite alphabet (52 characters if you count only lower-case and upper-case letters, or more if you count special symbols – either way, the alphabet is finite).

We will now enumerate the strings in such a way that each string appears exactly once in the list. We will use the same trick as used in Lecture note 10 to enumerate the elements of $\{0,1\}^*$ We get our bijection by setting $f(n)$ to be the n-th string in the list. List all strings of length 1 in lexicographic order, and then all strings of length 2 in lexicographic order, and then strings of length 3 in lexicographic order, and so forth. Since at each step, there are only finitely many strings of a particular length ℓ, any string of finite length appears in the list. It is also clear that each string appears exactly once in this list.

(c) No, the strings are still countable. Let $\mathscr{A} = \{a_1, a_2, \dots\}$ denote the alphabet. (We are making use of the fact that the alphabet is countably infinite when we assume there is such an enumeration.) We will provide two solutions:

Alternative 1: We will enumerate all the strings similar to that in part (b), although the enumeration requires a little more finesse. Notice that if we tried to list all strings of length 1, we would be stuck forever, since the alphabet is infinite! On the other hand, if we try to restrict our alphabet and only print out strings containing the first character $a \in \mathscr{A}$, we would also have a similar problem: the list

$$a, aa, aaa, \dots$$

also does not end.

The idea is to restrict *both* the length of the string and the characters we are allowed to use:

(a) List all strings containing only a_1 which are of length at most 1.

(b) List all strings containing only characters in $\{a_1, a_2\}$ which are of length at most 2 and have not been listed before.

(c) List all strings containing only characters in $\{a_1, a_2, a_3\}$ which are of length at most 3 and have not been listed before.

(d) Proceed onwards.

At each step, we have restricted ourselves to a finite alphabet with a finite length, so each step is guaranteed to terminate. To show that the enumeration is complete, consider any string s of length ℓ; since the length is finite, it can contain at most ℓ distinct a_i from the alphabet. Let k denote the largest index of any a_i which appears in s. Then, s will be listed in step $\max(k, \ell)$, so it appears in the enumeration. Further, since we are listing only those strings that have not appeared before, each string appears exactly once in the listing.

Alternative 2: We will encode the strings into ternary strings. Recall that we used a similar trick in Lecture note 10 to show that the set of all polynomials with natural coefficients is

countable. Suppose, for example, we have a string: $S = a_5a_2a_7a_4a_6$. Corresponding to each of the characters in this string, we can write its index as a binary string: $(101, 10, 111, 100, 110)$. Now, we can construct a ternary string where "2" is inserted as a separator between each binary string. Thus we map the string S to a ternary string: 101210211121002110. It is clear that this mapping is injective, since the original string S can be uniquely recovered from this ternary string. Thus we have an injective map to $\{0,1,2\}^*$. From Lecture note 10, we know that the set $\{0,1,2\}^*$ is countable, and hence the set of all strings with finite length over $\mathscr{A}$ is countable.

3 Python Functions

(a) The set $F = \{f : \mathbb{N} \to \{0,1\}\}$ is not countable.

(b) Prove that the set of all python functions that output $\{0,1\}$ is countable. (Python functions have the same power as Turing machines, but people are more familiar with python.)

(c) The set of Python functions that take in input x and output either 0 or 1 appears to be the same as F in (a), but the set of Python function is countable. Why?

Solution:

(a) If F is countable, then there is a list of functions f_i that $F = \{f_i, i \in \mathbb{N}\}$, however $f'(x) = 1 - f_x(x)$ is not in F, contradiction!

(b) A python function can be encoded into a binary string. There is a one-on-one mapping between binary string and integers in the following way:

- A finite binary string can be mapped to a natural number by taking it's binary form.
- That is, a binary string S of length n can be converted into integer $\sum_{i=0}^{n-1} S_i 2^i$.

So there is a bijection between set of finite length binary strings and the natural numbers.

(c) This is a bit subtle. In part (a), we used the fact that a function $f : \mathbb{N} \to 0,1$ assigns a value to each element of $\mathbb{N}$. While a python program may do the same thing, that is, each python program may assign a value to every elemengt of $\mathbb{N}$, python programs cannot produce all functions. Indeed, the number of functions that python programs can produce cannot be more than the number of python programs. This is countable by previous part. (Intuitively, a python program cannot contain more "information" than its length, whereas a function f contains some information for every natural number which means it contains infinite information. This is basically the same argument.)

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 8B

1 Countability and the Halting Problem

Prove the Halting Problem using the set of all programs and inputs.

a) What is a reasonable representation for a computer program? Using this definition, show that the set of all programs are countable. *(Hint: Python Code)*

b) We consider only finite-length inputs. Show that the set of all inputs are countable.

c) Assume that you have a program that tells you whether or not a given program halts on a specific input. Since the set of all programs and the set of all inputs are countable, we can enumerate them and construct the following table.

	x_1	x_2	x_3	x_4	...
p_1	H	L	H	L	...
p_2	L	L	L	H	...
p_3	H	L	H	L	...
p_4	L	H	L	L	...
⋮	⋮	⋮	⋮	⋮	⋱

An H (resp. L) in the ith row and jth column means that program p_i halts (resp. loops) on input x_j. Now write a program that is not within the set of programs in the table above.

d) Find a contradiction in part a and part c to show that the halting problem can't be solved.

Solution:

a) As in discussion and lecture, we represent a computer programs with a set of finite-length strings (which, int turn, can be represented by a set of finite length binary strings). The set of finite length binary strings are countably infinite. Therefore the set of all programs is countable.

b) Notice that all inputs can also be represented by a set of finite length binary strings. The set of finite length binary strings are countably infinite, as proved in Note 11. Therefore the set of all inputs is countable.

c) For the sake of deriving a contradiction in part (d), we will use the following program:

```
procedure P'(x_j)
    if P_j(x_j) halts then
        loop
```

```
        else
            halt
        end if
    end procedure
```

d) If the program you wrote in part c) exists, it must occur somewhere in our complete list of programs, P_n. This cannot be. Say that P_n has source code x_j (i.e. its source code corresponds to column j). What is the (i, j)th entry of the table? If it's H, then $P_n(x_j)$ should loop forever, by construction; if it's L, then $P_n(x_j)$ should halt. In either case, we have a contradiction.

2 Fixed Points

Consider the problem of determining if a function F has any fixed points. That is, given a function F that takes inputs from some (possibly infinite) set $\mathscr{X}$, we want to know if there is any input $x \in \mathscr{X}$ such that $F(x)$ outputs x. Prove that this problem is undecidable.

Solution:

We can prove this by reducing from the Halting Problem. Suppose we had some function `FixedPoint`(F) that solved the fixed-point problem. That is, we supply a `FixedPoint` a function F, and it outputs `true` if it can find some $x \in \mathscr{X}$ such that $F(x)$ outputs x, and `false` if no such x exists. We can define `TestHalt`(F,x) as follows:

```
def TestHalt(F, x):
  def F_prime(y):
    F(x)
    return y
  return FixedPoint(F_prime)
```

If $F(x)$ halts, we have that $F'(y)$ will always just return y, so every input is a fixed point. On the other hand, if $F(x)$ does not halt, F' won't return anything for any input y, so there can't be any fixed points. Thus, our definition of `TestHalt` must always work, which is a contradiction; this tells us that `FixedPoint` cannot exist.

3 Computability

Decide whether the following statements are true or false. Please justify your answers.

(a) The problem of determining whether a program halts in time 2^{n^2} on an input of size n is undecidable.

(b) There is no computer program `Line` which takes a program P, an input x, and a line number L, and determines whether the L^{th} line of code is executed when the program P is run on the input x.

Solution:

(a) False. You can simulate a program for 2^{n^2} steps and see if it halts.

Generally, we can always run a program for any fixed *finite* amount of time to see what it does. The problem of undecidability arises when no bounds on time are available.

(b) True.

We implement `Halt` which takes a program P, an input x and decides whether $P(x)$ halts, using `Line` as follows. We take the input P and modify it so that each exit or return statement jumps to a particular new line. Call the resulting program P'. We then hand that program to `Line` along with the input x and the number of the new line. If the original program halts than `Line` would return true, and if not `Line` would return false.

This contradicts the fact that the program `Halt` does not exist, so `Line` does not exist either.

At a high level, you can show the undecidability of a problem by using your program which solves the problem as a subroutine to solve a different problem that we know is undecidable. Alternatively, you can do a diagonalization proof like we did for `Halt`. The first approach is natural for computer programmers and flows from the fact that you are given P as text! Therefore you can look at it and modify it. This is what the solution above does.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 9A

1 Venn Diagram

Out of 1,000 computer science students, 400 belong to a club (and may work part time), 500 work part time (and may belong to a club), and 50 belong to a club and work part time.

(a) Suppose we choose a student uniformly at random. Let C be the event that the student belongs to a club and P the event that the student works part time. Draw a picture of the sample space Ω and the events C and P.

(b) What is the probability that the student belongs to a club?

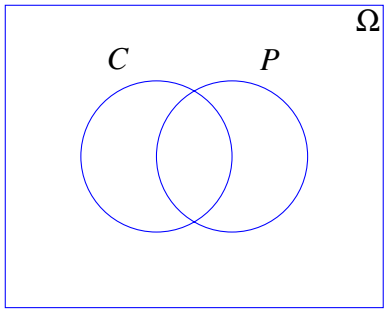
(c) What is the probability that the student works part time?

(d) What is the probability that the student belongs to a club AND works part time?

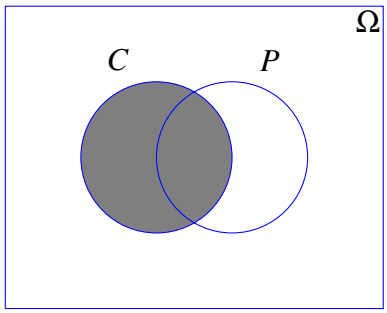
(e) What is the probability that the student belongs to a club OR works part time?

Solution:

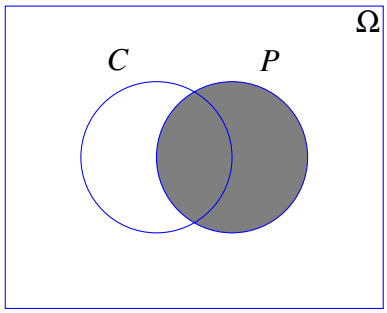
(a) The sample space will be illustrated by a Venn diagram.



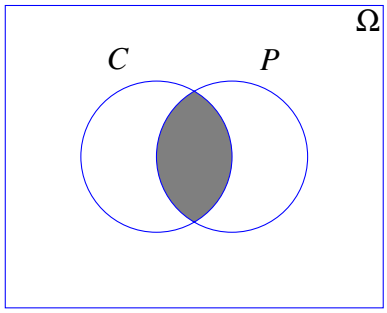
(b) $\mathbb{P}[C] = \frac{|C|}{|\Omega|} = \frac{400}{1000} = \frac{2}{5}$.



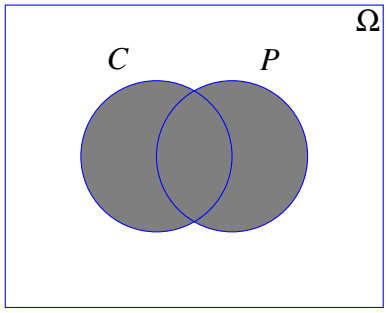
(c) $\mathbb{P}[P] = \dfrac{|P|}{|\Omega|} = \dfrac{500}{1000} = \dfrac{1}{2}$.



(d) $\mathbb{P}[P \cap C] = \dfrac{|P \cap C|}{|\Omega|} = \dfrac{50}{1000} = \dfrac{1}{20}$.



(e) $\mathbb{P}[P \cup C] = \mathbb{P}[P] + \mathbb{P}[C] - \mathbb{P}[P \cap C] = \dfrac{1}{2} + \dfrac{2}{5} - \dfrac{1}{20} = \dfrac{17}{20}$.



2 Flippin' Coins

Suppose we have an unbiased coin, with outcomes H and T, with probability of heads $\mathbb{P}[H] = 1/2$ and probability of tails also $\mathbb{P}[T] = 1/2$. Suppose we perform an experiment in which we toss the coin 3 times. An outcome of this experiment is (X_1, X_2, X_3), where $X_i \in \{H, T\}$.

(a) What is the *sample space* for our experiment?

(b) Which of the following are examples of *events*? Select all that apply.

- $\{(H,H,T),(H,H),(T)\}$
- $\{(T,H,H),(H,T,H),(H,H,T),(H,H,H)\}$
- $\{(T,T,T)\}$
- $\{(T,T,T),(H,H,H)\}$
- $\{(T,H,T),(H,H,T)\}$

(c) What is the complement of the event $\{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,T,T)\}$?

(d) Let A be the event that our outcome has 0 heads. Let B be the event that our outcome has exactly 2 heads. What is $A \cup B$?

(e) What is the probability of the outcome (H,H,T)?

(f) What is the probability of the event that our outcome has exactly two heads?

(g) What is the probability of the event that our outcome has at least one head?

Solution:

(a) $\Omega = \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}$

(b) An event must be a subset of Ω, meaning that it must consist of possible outcomes.

- No
- Yes
- Yes
- Yes
- Yes

(c) $\{(T,H,H),(T,H,T),(T,T,H)\}$

(d) $\{(T,H,H),(H,H,T),(H,T,H),(T,T,T)\}$

(e) Since $|\Omega| = 2^3 = 8$ and every outcome has equal probability, $\mathbb{P}[(H,H,T)] = 1/8$.

(f) The event of interest is $E = \{(H,H,T),(H,T,H),(T,H,H)\}$, which has size 3. Whence $\mathbb{P}[E] = 3/8$.

(g) If we do not see at least one head, then we must see at exactly three tails. The event $\overline{E} = \{(T,T,T)\}$ of seeing exactly three tails is thus the complement of the event E that we see at least one head. $\overline{E}$ occurs with probability $(1/2)^3 = 1/8$, so its complement E must occur with probability $1 - 1/8 = 7/8$.

3 Counting & Probability

Consider the equation $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 70$, where each x_i is a non-negative integer. We choose one of these solutions uniformly at random.

(a) What is the size of the sample space?

(b) What is the probability that both $x_1 \geq 30$ and $x_2 \geq 30$?

(c) What it the probability that either $x_1 \geq 30$ or $x_2 \geq 30$?

Solution:

(a) $\binom{75}{5}$. This is stars and bars.

(b) Put 30 balls each into the x_1 bin and the x_2 bin. We are left with 10 balls to distribute, whence there are $\binom{15}{5}$ possibilities. So the probability is $\binom{15}{5}/\binom{75}{5}$.

(c) Let A_i be the event that $x_i \geq 30$, then by inclusion-exclusion $\mathbb{P}[A_1 \cup A_2] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] = \left[\binom{45}{5} + \binom{45}{5} - \binom{15}{5}\right] / \binom{75}{5}$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 9B

1 Box of Marbles

You are given two boxes: one of them containing 900 red marbles and 100 blue marbles, the other one contains 500 red marbles and 500 blue marbles.

(a) If we pick one of the boxes randomly, and pick a marble what is the probability that it is blue?

(b) If we see that the marble is blue, what is the probability that it is chosen from box 1?

(c) Suppose we pick one marble from box 1 and without looking at its color we put it aside. Then we pick another marble from box 1. What is the probability that the second marble is blue?

Solution:

(a) Let B be the event that the picked marble is blue, R be the event that it is red, A_1 be the event that the marble is picked from box 1, and A_2 be the event that the marble is picked from box 2. Therefore we want to calculate $\mathbb{P}(B)$. By total probability,

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1)P(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2) = 0.5 \times 0.1 + 0.5 \times 0.5 = 0.3.$$

(b) In this part, we want to find $\mathbb{P}(A_1 \mid B)$. By Bayes rule,

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2)} = \frac{0.1 \times 0.5}{0.5 \times 0.1 + 0.5 \times 0.5} = \frac{1}{6}.$$

(c) Let B_1 be the event that first marble is blue, R_1 be the event that the first marble is red, and B_2 be the event that second marble is blue without looking at the color of first marble. We want to find $\mathbb{P}(B_2)$. By total probability,

$$\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(B_2 \mid R_1)\mathbb{P}(R_1) = \frac{99}{999} \times 0.1 + \frac{100}{999} \times 0.9 = 0.1.$$

More generally, one can see that the probability that the n-th marble picked from box 1 is blue with probability 0.1. This is clear by symmetry: all the permutations of the 1000 marbles have the same probability, so the probability that the n-th marble is blue is the same as the probability that the first marble is blue.

2 Duelling Meteorologists

Tom is a meteorologist in New York. On days when it snows, Tom correctly predicts the snow 70% of the time. When it doesn't snow, he correctly predicts no snow 95% of the time. In New York, it snows on 10% of all days.

(a) If Tom says that it is going to snow, what is the probability it will actually snow?

(b) Let A be the event that, on a given day, Tom predicts the weather correctly. What is $\mathbb{P}(A)$?

(c) Tom's friend Jerry is a meteorologist in Alaska. Jerry claims that she is a better meteorologist than Tom even though her overall accuracy is lower. After looking at their records, you determine that Jerry is indeed better than Tom at predicting snow on snowy days and sun on sunny day. Give an instance of the situation described above. *Hint: what is the weather like in Alaska?*

Solution:

(a) Let S be the event that it snows and T be the event that Tom predicts snow.

$$\begin{aligned}\mathbb{P}(S|T) &= \frac{\mathbb{P}(S\cap T)}{\mathbb{P}(T)} \\ &= \frac{\mathbb{P}(S)\cdot\mathbb{P}(T|S)}{\mathbb{P}(S\cap T)+\mathbb{P}(\overline{S}\cap T)} \\ &= \frac{\frac{1}{10}\times\frac{7}{10}}{\frac{1}{10}\times\frac{7}{10}+\frac{9}{10}\times\frac{5}{100}} = \frac{14}{23}\end{aligned}$$

(b)

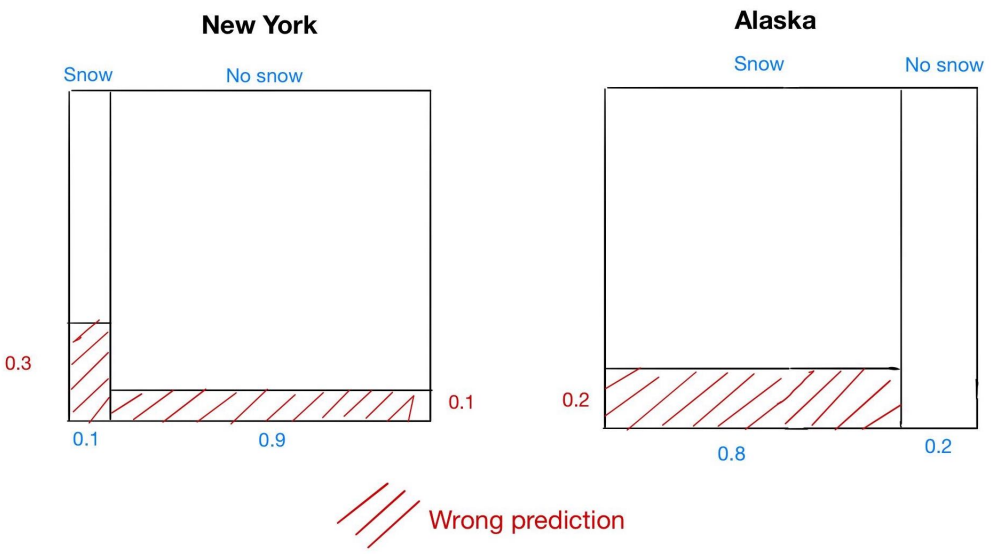
$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(S\cap T)+\mathbb{P}(\overline{S}\cap\overline{T}) \\ &= \frac{1}{10}\times\frac{7}{10}+\frac{9}{10}\times\frac{95}{100} = \frac{37}{40}\end{aligned}$$

(c) Even though Jerry's overall accuracy is lower, it is still possible that she is a better meteorologist if the weather is different.

For example, let's assume that it snows 80% of days in Alaska.

- When it snows, Jerry correctly predicts snow 80% of the time.
- When it doesn't snow, Jerry correctly predicts no snow 100% of the time.

Jerry's overall accuracy turns out to be less than Tom's even though she is better at predicting both categories! The following diagram gives an illustration of the situation. The intuition is that Jerry's error gets penalized more heavily than Tom because it snows more often in Alaska.



For more info on this kind of phenomena, check out Simpson's Paradox!

3 Binary Conditional Probabilities

Let us consider a sample space $\Omega = \{\omega_1, \ldots, \omega_N\}$ of size $N > 2$, and two probability functions $\mathbb{P}_1$ and $\mathbb{P}_2$ on it. That is, we have two probability spaces: $(\Omega, \mathbb{P}_1)$ and $(\Omega, \mathbb{P}_2)$.

If for every subset $A \subset \Omega$ of size $|A| = 2$ and every outcome $\omega \in \Omega$ it is true that $\mathbb{P}_1(\omega \mid A) = \mathbb{P}_2(\omega \mid A)$, then is it necessarily true that $\mathbb{P}_1(\omega) = \mathbb{P}_2(\omega)$ for all $\omega \in \Omega$? That is, if $\mathbb{P}_1$ and $\mathbb{P}_2$ are equal conditional on events of size 2, are they equal unconditionally? (*Hint*: Remember that probabilities must add up to 1.)

Solution: Yes, this is indeed true. To see why, let's take the subset $A = \{\omega_i, \omega_j\}$ for some $i, j \in \{1, \ldots, N\}$ and compute: For any $k \in \{1, 2\}$, we have $\mathbb{P}_k(\omega_i \mid A) = \frac{\mathbb{P}_k(\omega_i)}{\mathbb{P}_k(A)}$. Since this expression (by assumption) is the same for $k = 1$ and $k = 2$, we conclude that $\frac{\mathbb{P}_1(\omega_i)}{\mathbb{P}_2(\omega_i)} = \frac{\mathbb{P}_1(A)}{\mathbb{P}_2(A)}$. Repeating the reasoning for ω_j, we similarly find that $\frac{\mathbb{P}_1(\omega_j)}{\mathbb{P}_2(\omega_j)} = \frac{\mathbb{P}_1(A)}{\mathbb{P}_2(A)}$, and whence $\frac{\mathbb{P}_1(\omega_i)}{\mathbb{P}_1(\omega_j)} = \frac{\mathbb{P}_2(\omega_i)}{\mathbb{P}_2(\omega_j)}$. Since this is true for any $i, j \in \{1, \ldots, N\}$, we can sum over i to get

$$\frac{1}{\mathbb{P}_1(\omega_j)} = \sum_{i=1}^{N} \frac{\mathbb{P}_1(\omega_i)}{\mathbb{P}_1(\omega_j)} = \sum_{i=1}^{N} \frac{\mathbb{P}_2(\omega_i)}{\mathbb{P}_2(\omega_j)} = \frac{1}{\mathbb{P}_2(\omega_j)},$$

which shows that $\mathbb{P}_1(\omega_j) = \mathbb{P}_2(\omega_j)$ for all $j \in \{1, \ldots, N\}$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 10A

1 Probability Potpourri

Prove a brief justification for each part.

(a) For two events A and B in any probability space, show that $\mathbb{P}(A \setminus B) \geq \mathbb{P}(A) - \mathbb{P}(B)$.

(b) If $|\Omega| = n$, how many distinct events does the probability space have?

(c) Suppose $\mathbb{P}(D \mid C) = \mathbb{P}(D \mid \overline{C})$, where $\overline{C}$ is the complement of C. Prove that D is independent of C.

Solution:

(a) Start with the right side:

$$\begin{aligned}\mathbb{P}(A) - \mathbb{P}(B) &= [\mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B)] - [\mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)] \\ &= \mathbb{P}(A \setminus B) - \mathbb{P}(B \setminus A) \\ &\leq \mathbb{P}(A \setminus B)\end{aligned}$$

(b) An event is a subset of Ω, and for each outcome, there are 2 options: the outcome is in the event, or it isn't. Since there are n outcomes, there are 2^n events.

(c) Using total probability rule:

$$\mathbb{P}(D) = \mathbb{P}(D \cap C) + \mathbb{P}(D \cap \overline{C}) = \mathbb{P}(D \mid C) \cdot \mathbb{P}(C) + \mathbb{P}(D \mid \overline{C}) \cdot \mathbb{P}(\overline{C})$$

But we know that $\mathbb{P}(D \mid C) = \mathbb{P}(D \mid \overline{C})$, so this simplifies to

$$\mathbb{P}(D) = \mathbb{P}(D \mid C) \cdot [\mathbb{P}(C) + \mathbb{P}(\overline{C})] = \mathbb{P}(D \mid C) \cdot 1 = \mathbb{P}(D \mid C)$$

which defines independence.

2 Aces

Consider a standard 52-card deck of cards:

(a) Find the probability of getting an ace or a red card, when drawing a single card.

(b) Find the probability of getting an ace or a spade, but not both, when drawing a single card.

(c) Find the probability of getting the ace of diamonds when drawing a 5 card hand.

(d) Find the probability of getting exactly 2 aces when drawing a 5 card hand.

(e) Find the probability of getting at least 1 ace when drawing a 5 card hand.

(f) Find the probability of getting at least 1 ace or at least 1 heart when drawing a 5 card hand.

Solution:

(a) Inclusion-Exclusion Principle: $\frac{4}{52}+\frac{26}{52}-\frac{2}{52}=\frac{28}{52}=\frac{7}{13}$.

(b) Inclusion-Exclusion, but we exclude the intersection: $\frac{4}{52}+\frac{13}{52}-2\cdot\frac{1}{52}=\frac{15}{52}$.

(c) Ace of diamonds is fixed, but the other 4 cards in the hand can be any other card: $\frac{\binom{51}{4}}{\binom{52}{5}}$.

(d) Account for the number of ways to draw 2 aces and the number of ways to draw 3 non-aces: $\frac{\binom{4}{2}\cdot\binom{48}{3}}{\binom{52}{5}}$.

(e) Complement to getting no aces: $\mathbb{P}[\text{at least one ace}]=1-\mathbb{P}[\text{zero aces}]=1-\frac{\binom{48}{5}}{\binom{52}{5}}$.

(f) Complement to getting no aces and no hearts: $\mathbb{P}[\text{at least one ace OR at least one heart}]=1-\mathbb{P}[\text{zero aces AND zero hearts}]=1-\frac{\binom{36}{5}}{\binom{52}{5}}$. This is because $52-13-3=36$, where 13 is the number of hearts and 3 is the number of non-heart aces.

3 Balls and Bins

Throw n balls into n labeled bins one at a time.

(a) What is the probability that the first bin is empty?

(b) What is the probability that the first k bins are empty?

(c) Let A be the event that at least k bins are empty. Notice that there are $m=\binom{n}{k}$ sets of k bins out of the total n bins. If we assune A_i is the event that the i^{th} set of k bins is empty. Then we can write A as the union of A_i's.

$$A=\bigcup_{i=1}^{m}A_i.$$

Write the union bound for the probability A.

(d) Use the union bound to give an upper bound on the probability A from part (c).

(e) What is the probability that the second bin is empty given that the first one is empty?

(f) Are the events that "the first bin is empty" and "the first two bins are empty" independent?

(g) Are the events that "the first bin is empty" and "the second bin is empty" independent?

Solution: Since the balls are thrown one at a time, there is an ordering, and so we are sampling with replacement where order matters rather than where it doesn't (which would correspond to each configuration in the stars and bars setup being equally likely).

(a) The probability that ball i does not land in the first bin is $\frac{n-1}{n}$. The probability that all of the balls do not land in the first bin is $\left(\frac{n-1}{n}\right)^n$.

(b) The probability that ball i does not land in the first k bins is $\frac{n-k}{n}$. The probability that all of the balls do not land in the first k bins is $\left(\frac{n-k}{n}\right)^n$.

(c) We use the union bound. Then

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{m} A_i\right) \leq \sum_{i=1}^{m} \mathbb{P}(A_i)$$

(d) We know the probability of the first k bins being empty from part (b), and this is true for any set of k bins, so

$$\mathbb{P}(A_i) = \left(\frac{n-k}{n}\right)^n.$$

Then,

$$\mathbb{P}(A) \leq m \cdot \left(\frac{n-k}{n}\right)^n = \binom{n}{k}\left(\frac{n-k}{n}\right)^n.$$

(e) Using Bayes' Rule:

$$\begin{aligned} \mathbb{P}[\text{2nd bin empty} \mid \text{1st bin empty}] &= \frac{\mathbb{P}[\text{2nd bin empty} \cap \text{1st bin empty}]}{\mathbb{P}[\text{1st bin empty}]} \\ &= \frac{(n-2)^n/n^n}{(n-1)^n/n^n} \\ &= \left(\frac{n-2}{n-1}\right)^n \end{aligned} \tag{1}$$

Alternate solution:
We know bin 1 is empty, so each ball that we throw can land in one of the remaining $n-1$ bins. We want the probability that bin 2 is empty, which means that each ball cannot land in bin 2 either, leaving $n-2$ bins. Thus for each ball, the probability that bin 2 is empty given that bin 1 is empty is $(n-2)/(n-1)$. For n total balls, this probability is $[(n-2)/(n-1)]^n$.

(f) They are dependent. Knowing the latter means the former happens with probability 1.

(g) In part (c) we calculated the probability that the second bin is empty given that the first bin is empty: $[(n-2)/(n-1)]^n$. The probability that the second bin is empty (without any prior information) is $[(n-1)/n]^n$. Since these probabilities are not equal, the events are dependent.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 10B

1 Pullout Balls

Suppose you have a bag containing six balls numbered $1, 2, 3, 4, 5, 6$.

(a) You perform the following experiment: pull out a single ball and record its number. What is the expected value of the number that you record?

(b) You repeat the experiment from part (a), except this time you pull out two balls together and record the product of their numbers. What is the expected value of the total that you record?

Solution:

(a) Let X be the number that you record. Each ball is equally likely to be chosen, so

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}(X = x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

As demonstrated here, the expected value of a random variable need not, and often is not, a feasible value of that random variable (there is no outcome ω for which $X(\omega) = 3.5$).

(b) Let Y be the product of two numbers that you pull out. Then

$$\mathbb{E}[Y] = \frac{\sum_{i=1}^{6} (i \times \sum_{j=i+1}^{6} j)}{\binom{6}{2}} = \frac{20 + 36 + 45 + 44 + 30}{15} = \frac{35}{3}$$

2 How Many Queens?

You shuffle a standard 52-card deck, before drawing the first three cards from the top of the pile. Let X denote the number of queens you draw.

(a) What is $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$, $\mathbb{P}(X = 2)$ and $\mathbb{P}(X = 3)$?

(b) What do your answers you computed in part a add up to?

(c) Compute $\mathbb{E}(X)$ from the definition of expectation.

(d) Are the X_i indicators independent?

Solution:

(a) Calculate each case of X = 0, 1, 2, 3:

We must draw 3 non-queen cards in a row, so the probability is

$$\mathbb{P}(X=0) = \frac{48}{52}\cdot\frac{47}{51}\cdot\frac{46}{50} = \frac{4324}{5525}.$$

Alternatively, every 3-card hand is equally likely, so we can use counting. There are $\binom{52}{3}$ total 3-card hands, and $\binom{48}{3}$ hands with only non-queen cards, which gives us the same result.

$$\mathbb{P}(X=0) = \frac{\binom{48}{3}}{\binom{52}{3}} = \frac{4324}{5525}$$

- We will continue to use counting. The number of hands with exactly one queen amounts to the number of ways to choose 1 queen out of 4, and 2 non-queens out of 48.

$$\mathbb{P}(X=1) = \frac{\binom{4}{1}\binom{48}{2}}{\binom{52}{3}} = \frac{1128}{5525}$$

- Choose 2 queens out of 4, and 1 non-queen out of 48.

$$\mathbb{P}(X=2) = \frac{\binom{4}{2}\binom{48}{1}}{\binom{52}{3}} = \frac{72}{5525}$$

- Choose 3 queens out of 4.

$$\mathbb{P}(X=3) = \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{1}{5525}$$

(b) We check:

$$\mathbb{P}(X=0)+\mathbb{P}(X=1)+\mathbb{P}(X=2)+\mathbb{P}(X=3) = \frac{4324+1128+72+1}{5525} = 1$$

(c) From the definition, $\mathbb{E}(X) = \sum_{k=0}^{3} k\mathbb{P}(X=k)$, so

$$\mathbb{E}(X) = 0\cdot\frac{4324}{5525} + 1\cdot\frac{1128}{5525} + 2\cdot\frac{72}{5525} + 3\cdot\frac{1}{5525} = \frac{3}{13}.$$

(d) No, they are not independent. As an example:

$$\mathbb{P}(X_1=1)\mathbb{P}(X_2=1) = \frac{1}{13}\cdot\frac{1}{13} = \frac{1}{169}$$

However,

$$\mathbb{P}(X_1=1, X_2=1) = \mathbb{P}(\text{the first and second cards are both queens}) = \frac{4}{52}\cdot\frac{3}{51} = \frac{1}{221}.$$

3 Head Count

Consider a coin with $\mathbb{P}(\text{Heads}) = 2/5$. Suppose you flip the coin 20 times, and define X to be the number of heads.

(a) Name the distribution of X and what its parameters are.

(b) What is $\mathbb{P}(X = 7)$?

(c) What is $\mathbb{P}(X \geq 1)$? Hint: You should be able to do this without a summation.

(d) What is $\mathbb{P}(12 \leq X \leq 14)$?

Solution:

(a) Since we have 20 independent trials, with each trial having a probability $2/5$ of success, $X \sim \text{Binomial}(20, 2/5)$.

(b)

$$\mathbb{P}(X = 7) = \binom{20}{7}\left(\frac{2}{5}\right)^7\left(\frac{3}{5}\right)^{13}.$$

(c)

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \left(\frac{3}{5}\right)^{20}.$$

(d)

$$\begin{aligned}\mathbb{P}(12 \leq X \leq 14) &= \mathbb{P}(X = 12) + \mathbb{P}(X = 13) + \mathbb{P}(X = 14) \\ &= \binom{20}{12}\left(\frac{2}{5}\right)^{12}\left(\frac{3}{5}\right)^8 + \binom{20}{13}\left(\frac{2}{5}\right)^{13}\left(\frac{3}{5}\right)^7 + \binom{20}{14}\left(\frac{2}{5}\right)^{14}\left(\frac{3}{5}\right)^6.\end{aligned}$$

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 11A

1 Linearity

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears?

Solution:

(a) Let A_i be the indicator you win the ith time you play game A and B_i be the same for game B. The expected value of A_i and B_i are

$$\mathbb{E}[A_i] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3},$$
$$\mathbb{E}[B_i] = 1 \cdot \frac{1}{5} + 0 \cdot \frac{4}{5} = \frac{1}{5}.$$

Let T_A be the random variable for the number of tickets you win in game A, and T_B be the number of tickets you win in game B.

$$\begin{aligned}\mathbb{E}[T_A + T_B] &= 3\,\mathbb{E}[A_1] + \cdots + 3\,\mathbb{E}[A_{10}] + 4\,\mathbb{E}[B_1] + \cdots + 4\,\mathbb{E}[B_{20}] \\ &= 10\left(3 \cdot \frac{1}{3}\right) + 20\left(4 \cdot \frac{1}{5}\right) = 26\end{aligned}$$

(b) There are $1,000,000 - 4 + 1 = 999,997$ places where "book" can appear, each with a (non-independent) probability of $1/26^4$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1 + \cdots + A_{999,997}] \\ &= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_{999,997}] \\ &= \frac{999,997}{26^4} \approx 2.19.\end{aligned}$$

2 Joint Distributions

(a) Give an example of discrete random variables X and Y with the property that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. You should specify the joint distribution of X and Y.

(b) Give an example of discrete random variables X and Y that (i) are *not independent* and (ii) have the property that $\mathbb{E}[XY] = 0$, $\mathbb{E}[X] = 0$, and $\mathbb{E}[Y] = 0$. Again you should specify the joint distribution of X and Y.

Solution:

(a) Let $P(X = 1) = \frac{1}{2}$, $P(X = -1) = \frac{1}{2}$, and $Y \equiv X$. Then $\mathbb{E}[X] = 1\mathbb{P}[X = 1] + (-1)\mathbb{P}[X = -1] = 0$, and $\mathbb{E}[Y] = \mathbb{E}[X]$. Similarly, since $X = Y$, $\mathbb{E}[XY] = \mathbb{E}[X^2] = 1$ and $\mathbb{E}[X]\mathbb{E}[Y] = 0$.

(b) One example is given by $P(X = -1, Y = \frac{1}{3}) = P(X = 1, Y = \frac{1}{3}) = P(X = 0, Y = -\frac{2}{3}) = \frac{1}{3}$.

3 Ball in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

(b) What is the expected number of empty bins?

(c) Define a collision to occur when two balls land in the same bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?

Solution:

(a) We will use linearity of expectation. Note that the expectation of an indicator variable is just the probability the indicator variable = 1. (Verify for yourself that is true).

$$\mathbb{E}[X_i] = \mathbb{P}[\text{ball 1 falls into bin } i] + \mathbb{P}[\text{ball 2 falls into bin } i] \cdots = \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.$$

(b) Let X_i be the indicator variable denoting whether bin i ends up empty. This can happen if and only if all the balls end in the remaining $n-1$ bins, and this happens with a probability of $\left(\frac{n-1}{n}\right)^k$. Hence the expected number of empty bins is

$$\mathbb{E}[X_1 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = n\left(\frac{n-1}{n}\right)^k$$

(c) The number of collisions is the number of balls minus the number of occupied bins, since the first ball of every occupied bin is not a collision.

$$\begin{aligned}\mathbb{E}[\text{collisions}] &= k - \mathbb{E}[\text{occupied bins}] = k - n + \mathbb{E}[\text{empty locations}] \\ &= k - n + n\left(1 - \frac{1}{n}\right)^k\end{aligned}$$

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 11B

1 Variance Proofs

(a) Let X be a random variable. Prove that:

$$\text{Var}(X) \geq 0$$

(b) Let $X_1, \ldots, X_n$ be random variables. Prove that:

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$$

Hint: Without loss of generality we can assume that $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] = 0$. Why?

(c) Let $a_1, \ldots, a_n \in \mathbb{R}$, and $X_1, \ldots, X_n$ be random variables. Prove that:

$$\sum_{i=1}^{n} a_i^2 \cdot \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i \cdot a_j \cdot \text{cov}(X_i, X_j) \geq 0$$

Solution:

(a) By definition, we have that

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \sum_k (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) \end{aligned}$$

Each term in this summation is non-negative, since it is the product of a squared number and a probability, both of which are guaranteed to be non-negative. Since the sum of a bunch of non-negative terms must itself be non-negative, we have that $\text{Var}(X) \geq 0$.

(b) We first note that for a constant c, $\text{Var}(X + c) = \text{Var}(X)$ and similarly that $\text{cov}(X + c, Y) = \text{cov}(X, Y)$. Thus, we can subtract $\mathbb{E}[X_i]$ from X_i without changing anything in our target equality; this reduces us to the case where all the means are zero. Hence, we can write

$$\begin{aligned} \text{Var}(X_1 + \cdots + X_n) &= \mathbb{E}\left[(X_1 + \cdots + X_n)^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \end{aligned}$$

where we used the fact that the means were all zero in the first and last steps to simplify the definitions of variance and covariance.

(c) We start with the left hand side of our desired inequality. Factoring the constants into the variance and covariance, we get

$$\sum_{i=1}^{n} a_i^2 \cdot \text{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i \cdot a_j \cdot \text{cov}(X_i, X_j) = \sum_{i=1}^{n} \text{Var}(a_i X_i) + 2 \sum_{1 \le i < j \le n} \text{cov}(a_i X_i, a_j X_j)$$

By part (b), we have that

$$\sum_{i=1}^{n} \text{Var}(a_i X_i) + 2 \sum_{1 \le i < j \le n} \text{cov}(a_i X_i, a_j X_j) = \text{Var}(a_1 X_1 + \cdots + a_n X_n)$$

Since $a_1 X_1 + \ldots + a_n X_n$ is a discrete random variable, part (a) tells us that its variance is non-negative. Putting these all together, we get that

$$\sum_{i=1}^{n} a_i^2 \cdot \text{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i \cdot a_j \cdot \text{cov}(X_i, X_j) \ge 0$$

2 Subset Card Game

Jonathan and Yiming are playing a card game. Jonathan has $k > 2$ cards, and each card has a real number written on it. Jonathan tells Yiming (truthfully), that the sum of the card values is 0, and that the sum of squares of the values on the cards is 1. Specifically, if the card values are $c_1, c_2, \ldots, c_k$, then we have $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} c_i^2 = 1$.

The cards are then going to be dealt randomly in the following fashion: for each card in the deck, a fair coin is flipped. If the coin lands heads, then the card goes to Yiming, and if the coin lands tails, the card goes to Jonathan. Note that it is possible for either player to end up with no cards/all the cards.

Calculate $\text{Var}(S)$, where S is the sum of value of cards in Yiming's hand. The answer should not include a summation.

Solution: Let I_i be the indicator random variable indicating whether or not card i goes to Yiming. We have $S = \sum_{i=1}^{k} c_i I_i$ as the value of Yiming's hand. Then, we see that $\mathbb{E}[S] = \sum_{i=1}^{k} c_i \cdot \frac{1}{2} = 0$ and

$$\begin{aligned} \text{Var}(S) &= \sum_{i=1}^{k} \text{Var}(c_i I_i) \qquad \text{(due to independence) of } I_i \\ &= \sum_{i=1}^{k} c_i^2 \, \text{Var}(I_i) \end{aligned}$$

We know that I_i is a Bernoulli random variable, so its variance is $\frac{1}{4}$. Thus, we see that $\text{Var}(S) = \frac{1}{4}$.

3 Variance

A building has n upper floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each person gets off at one of the n upper floors uniformly at random and independently of everyone else. What is the *variance* of the number of floors the elevator *does not* stop at?

Solution: Let N be the number of floors the elevator does not stop at. We can represent N as the sum of the indicator variables $I_1, \ldots, I_n$, where $I_i = 1$ if no one gets off on floor i. Thus, we have

$$\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

$$\mathbb{E}[N] = \sum_{i=1}^{n} \mathbb{E}[I_i] = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, since $\text{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2$ the only piece we don't already know is $\mathbb{E}[N^2]$. We can calculate this by again expanding N as a sum:

$$\mathbb{E}[N^2] = \mathbb{E}[(I_1 + \cdots + I_n)^2] = \mathbb{E}\left[\sum_{i,j} I_i I_j\right] = \sum_{i,j} \mathbb{E}[I_i I_j] = \sum_{i}^{n} \mathbb{E}[I_i^2] + \sum_{i \neq j} \mathbb{E}[I_i I_j].$$

The first term is simple to calculate: since I_i is an indicator, $I_i^2 = I_i$, so we have

$$\mathbb{E}[I_i^2] == \mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

meaning that

$$\sum_{i=1}^{n} \mathbb{E}\left[I_i^2\right] = n\left(\frac{n-1}{n}\right)^m.$$

From the definition of the variables I_i, we see that $I_i I_j = 1$ when both I_i and I_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

$$\mathbb{P}[I_i = I_j = 1] = \mathbb{P}[I_i = 1 \cap I_j = 1] = \left(\frac{n-2}{n}\right)^m.$$

Thus we now know

$$\sum_{i \neq j} \mathbb{E}[I_i I_j] = n(n-1)\left(\frac{n-2}{n}\right)^m,$$

and we can assemble everything we've done so far to see that

$$\text{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - n^2\left(\frac{n-1}{n}\right)^{2m}.$$

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 12A

1 Inequality Practice

(a) X is a random variable such that $X > -5$ and $\mathbb{E}[X] = -3$. Find an upper bound for the probability of X being greater than or equal to -1.

(b) Y is a random variable such that $Y < 10$ and $\mathbb{E}[Y] = 1$. Find an upper bound for the probability of Y being less than or equal to -1.

(c) You roll a die 100 times. Let Z be the sum of the numbers that appear on the die throughout the 100 rolls. Compute $\text{Var}(Z)$. Then use Chebyshev's inequality to bound the probability of the sum Z being greater than 400 or less than 300.

Solution:

(a) We want to use Markov's Inequality, but recall that Markov's Inequality only works with non-negative random variables. So, we define a new random variable $\tilde{X} = X + 5$, where $\tilde{X}$ is always non-negative, so we can use Markov's on $\tilde{X}$. By linearity of expectation, $\mathbb{E}[\tilde{X}] = -3 + 5 = 2$. So, $\mathbb{P}[\tilde{X} \geq 4] \leq 2/4 = 1/2$.

(b) We again use Markov's Inequality. Similarly, define $\tilde{Y} = -Y + 10$, and $\mathbb{E}[\tilde{Y}] = -1 + 10 = 9$. $P[Y \leq -1] = P[-Y \geq 1] = P[-Y + 10 \geq 11] \leq 9/11$.

(c) Let Z_i be the number on the die for the ith roll, for $i = 1, \ldots, 100$. Then, $Z = \sum_{i=1}^{100} Z_i$. By linearity of expectation, $\mathbb{E}[Z] = \sum_{i=1}^{100} \mathbb{E}[Z_i]$.

$$\mathbb{E}[Z_i] = \sum_{j=1}^{6} j \cdot \mathbb{P}[Z_i = j] = \sum_{j=1}^{6} j \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j = \frac{1}{6} \cdot 21 = \frac{7}{2}$$

Then, we have $\mathbb{E}[Z] = 100 \cdot (7/2) = 350$.

$$\mathbb{E}[Z_i^2] = \sum_{j=1}^{6} j^2 \cdot \mathbb{P}[Z_i = j] = \sum_{j=1}^{6} j^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j^2 = \frac{1}{6} \cdot 91 = \frac{91}{6}$$

Then, we have

$$\text{Var}(Z_i) = \mathbb{E}[Z_i^2] - \mathbb{E}[Z_i]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

Since the Z_is are independent, and therefore uncorrelated, we can add the $\text{Var}(Z_i)$s to get $\text{Var}(Z) = 100(35/12)$.

Putting it all together, we use Chebyshev's to get

$$\mathbb{P}[|Z - 350| \geq 50] \leq \frac{100(35/12)}{50^2} = \frac{7}{60}.$$

2 Vegas

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip his or her coin. Assume that each person is independently likely to carry a fair or a trick coin.

1. Given the results of your experiment, how should you estimate p?
   (*Hint:* Construct an (unbiased) estimator for p such that $E[\hat{p}] = p$.)
2. How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

Solution:

1. We want to construct an estimate $\hat{p}$ such that $\mathbb{E}[\hat{p}] = p$. Then, if we have a large enough sample, we'd expect to get a good estimate of p. Let X_i be the indicator that the ith person's coin flips to a heads. What we observe is the fraction of people whose coin is heads. In other words, we measure $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. How can we use this observation to construct $\hat{p}$?

   First,
   $$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_i] = p\cdot 1 + (1-p)\cdot\frac{1}{2},$$
   where the last equality follows from total probability. Solving for p, we find that
   $$p = 2\,\mathbb{E}[X] - 1 = \mathbb{E}[2X-1].$$
   Thus, our estimator $\hat{p}$ should be $2X-1$.

2. We want to find n such that $P[|\hat{p}-p| \le 0.05] > 0.95$. Another way to state this is that we want
   $$P[|\hat{p}-p| > 0.05] \le 0.05.$$
   Notice that $\mathbb{E}[\hat{p}] = p$ by construction, so we can immediately apply Chebyshev's inequality on $\hat{p}$. What we get is:
   $$P[|\hat{p}-p| > 0.05] \le \frac{\mathrm{Var}[\hat{p}]}{0.05^2} \le 0.05$$
   So, we want n such that $\mathrm{Var}[\hat{p}] \le 0.05^3$.
   $$\mathrm{Var}[\hat{p}] = \mathrm{Var}[2X-1] = 4\,\mathrm{Var}[X] = \frac{4}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{4}{n}\mathrm{Var}[X_1].$$
   But X_i is an indicator (Bernoulli variable), so its variance is bounded by $\frac{1}{4}$ (note that $p(1-p)$ is maximized at $p=\frac{1}{2}$ to yield a value of $\frac{1}{4}$). Therefore we have
   $$\mathrm{Var}[\hat{p}] \le \frac{4}{n}\frac{1}{4} = \frac{1}{n}.$$
   So, we choose n such that $\frac{1}{n} \le 0.05^3$, so $n \ge \frac{1}{0.05^3} = 8000$.

3 Working with the Law of Large Numbers

(a) A fair coin is tossed multiple times and you win a prize if there are more than 60% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

(b) A fair coin is tossed multiple times and you win a prize if there are more than 40% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

(c) A fair coin is tossed multiple times and you win a prize if there are between 40% and 60% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

(d) A fair coin is tossed multiple times and you win a prize if there are exactly 50% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

Solution:

(a) 10 tosses. By LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being greater than 0.60 if there are 100 tosses (compared with 10 tosses).

(b) 100 tosses. Again, by LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being smaller than 0.40 if there are 100 tosses. A lower chance of being smaller than 0.40 is the desired result.

(c) 100 tosses. Again, by LLN, the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being both smaller than 0.40 if there are 100 tosses. Similarly, there is a lower chance of being larger than 0.60 if there are 100 tosses. Lower chances of both of these events is desired if we want the fraction of heads to be between 0.4 and 0.6.

(d) 10 tosses. Compare the probability of getting equal number of heads and tails between $2n$ and $2n+2$ tosses.

$$\begin{aligned}
\mathbb{P}[n \text{ heads in } 2n \text{ tosses}] &= \binom{2n}{n}\frac{1}{2^{2n}} \\
\mathbb{P}[n+1 \text{ heads in } 2n+2 \text{ tosses}] &= \binom{2n+2}{n+1}\frac{1}{2^{2n+2}} = \frac{(2n+2)!}{(n+1)!(n+1)!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{(2n+2)(2n+1)2n!}{(n+1)(n+1)n!n!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{2n+2}{n+1}\cdot\frac{2n+1}{n+1}\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} < \left(\frac{2n+2}{n+1}\right)^2\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} \\
&= 4\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} = \binom{2n}{n}\frac{1}{2^{2n}} = \mathbb{P}[n \text{ heads in } 2n \text{ tosses}]
\end{aligned}$$

As we increment n, the probability will always decrease. Therefore, the larger n is, the less probability we'll get exactly 50% heads. □

Note: By Stirling's approximation, $\binom{2n}{n}2^{-2n}$ is roughly $(\pi n)^{-1/2}$ for large n.

See https://github.com/dingyiming0427/CS70-demo/ for a code demo.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 12B

1 Planetary Party

(a) Suppose we are at party on a planet where every year is 2849 days. If 30 people attend this party, what is the exact probability that two people will share the same birthday? You may leave your answer as an unevaluated expression.

(b) From lecture, we know that given n bins and m balls, $\mathbb{P}[\text{no collision}] \approx \exp(-m^2/(2n))$. Using this, give an approximation for the probability in part (a).

(c) What is the minimum number of people that need to attend this party to ensure that the probability that any two people share a birthday is at least 0.5? You can use the approximation you used in the previous part.

(d) Now suppose that 70 people attend this party. What the is probability that none of these 70 individuals have the same birthday? You can use the approximation you used in the previous parts.

Solution:

(a) Let's compute the probability that no two partygoers have the same birthday. We know the second person at the party cannot share a birthday with the first person, the third person at the party cannot share a birthday with the first two, etc. Thus

$$\mathbb{P}[\text{no collision}] = \left(1 - \frac{1}{2849}\right)\left(1 - \frac{2}{2849}\right)\left(1 - \frac{3}{2849}\right)\cdots\left(1 - \frac{29}{2849}\right)$$

Thus $\mathbb{P}[\text{collision}] = 1 - \mathbb{P}[\text{no collision}] = 1 - \left(1 - \frac{1}{2849}\right)\left(1 - \frac{2}{2849}\right)\left(1 - \frac{3}{2849}\right)\cdots\left(1 - \frac{29}{2849}\right)$.

(b) From lecture, we know that given n bins and m balls, $\mathbb{P}[\text{no collision}] \approx \exp(-m^2/(2n))$. Therefore in this case, if we want to find the probability of collision, we must find $1 - \mathbb{P}[\text{no collision}]$.

$$\mathbb{P}[\text{no collision}] = \exp\left(-\frac{30^2}{2 \cdot 2849}\right) = 0.854$$

This means that there is a 0.146 chance that two people share the same birthday in the group of 30.

(c) Rephrasing the question in terms of balls and bins, we want to find the minimum number of balls (m) such that there is at least 0.5 probability of collision when we have $n = 2849$ bins, which is the same as at **most** 0.5 probability of **no** collisions.

$$\begin{aligned}\mathbb{P}[\text{no collisions}] &\approx \exp\left(\frac{-m^2}{n}\right) \leq 0.5 \\ &\implies \frac{-m^2}{n} \leq \ln 0.5 \\ &\implies m \geq \sqrt{(-2\ln 0.5)n} \\ &= 62.845\end{aligned}$$

Since m must be an integer which is at least 62.485, we need at least $\boxed{63}$ people at the party.

(d) Once again we need to find $\mathbb{P}[\text{no collisions}]$ given that $m = 70$.

$$\mathbb{P}[\text{no collision}] = \exp\left(-\frac{70^2}{2 \cdot 2849}\right) = 0.423$$

There is about a 42% chance that 70 people don't share the same birthday.

2 Throwing Balls into a Depth-Limited Bin

Say you want to throw n balls into n bins with depth $k-1$ (they can fit $k-1$ balls, after that the bins overflow). Suppose that n is a large number and $k = 0.1n$. You throw the balls randomly into the bins, but you would like it if they don't overflow. You feel that you might expect not too many balls to land in each bin, but you're not sure, so you decide to investigate the probability of a bin overflowing.

(a) Count the number of ways we can select k balls to put in the first bin, and then throw the remaining balls randomly. You should assume that the balls are distinguishable.

(b) Argue that your answer in (a) is an upper bound for the number of ways that the first bin can overflow.

(c) Calculate an upper bound on the probability that the first bin will overflow.

(d) Upper bound the probability that some bin will overflow. [*Hint*: Use the union bound.]

(e) How does the above probability scale as n gets really large?

Solution:

(a) We choose k of the balls to throw in the first bin and then throw the remaining $n-k$, giving us $\binom{n}{k} n^{n-k}$.

(b) Certainly any outcome of the ball-throwing that overflows the first bin is accounted for – we can simply choose the first k balls that land in the first bin and then simulate the rest of the outcome via random throwing. However, we are potentially overcounting: if $k+1$ balls go in the first bin, we have many choices for which k of them that could have been the "chosen" ones, and we count each one of these choices as distinct. However, they correspond to the same configuration, namely the one where $k+1$ balls are in the first bin. Hence we get an upper bound.

(c) We divide by the total number of ways the balls could have fallen into the bins, with order, so we get

$$\frac{\binom{n}{k}n^{n-k}}{n^n} = \frac{\binom{n}{k}}{n^k}.$$

(d) Let A_i denote the event that bin i overflows. By symmetry $\mathbb{P}(A_i) = \mathbb{P}(A_1)$ for all i. By the union bound we have

$$\mathbb{P}(\cup_i A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) \leq n\mathbb{P}(A_i) \leq n \cdot \frac{\binom{n}{k}}{n^k}.$$

(e) We get

$$n \cdot \frac{\binom{n}{k}}{n^k} = n \cdot \frac{n \cdot (n-1) \cdots (n-k+1)}{k!n^k} \leq n \cdot \frac{n^k}{k!n^k} = \frac{n}{k!} = \frac{n}{(0.1n) \cdot (k-1)!} = \frac{10}{(0.1n-1)!}.$$

Clearly, as n gets large this probability is going to 0. Note that this same analysis would work with $k = cn$ for any constant $0 < c < 1$. Hence, using some very coarse upper bounds, we can see that as the number of balls and bins grows, we have that it is very unlikely that we get a constant fraction of the balls in any single bin.

3 The Memoryless Property

Let X be a discrete random variable which takes on values in $\mathbb{Z}_+$. Suppose that for all $m, n \in \mathbb{N}$, we have $\mathbb{P}(X > m+n \mid X > n) = \mathbb{P}(X > m)$. Prove that X is a geometric distribution. Hint: In order to prove that X is geometric, it suffices to prove that there exists a $p \in [0,1]$ such that $\mathbb{P}(X > i) = (1-p)^i$ for all $i > 0$.

Solution:

Notice that

$$\mathbb{P}(X > m+n \mid X > n) = \frac{\mathbb{P}(X > m+n)}{\mathbb{P}(X > n)} = \mathbb{P}(X > m),$$

where the first equality holds from definition of conditional probability, and the second from the given in the question. So, this gives $\mathbb{P}(X > m+n) = \mathbb{P}(X > m)\mathbb{P}(X > n)$.

$$\mathbb{P}(X > m) = \mathbb{P}(X > m+n \mid X > n) = \frac{\mathbb{P}(X > m+n)}{\mathbb{P}(X > n)},$$

where that the first equality comes from the given in the question, and the second equality holds from definition of conditional. So, this gives $\mathbb{P}(X > m+n) = \mathbb{P}(X > m)\mathbb{P}(X > n)$.

By repeatedly applying this property, we can deduce $\mathbb{P}(X > n) = \mathbb{P}(X > 1+\cdots+1) = \mathbb{P}(X > 1)^n$. Let $p := 1 - \mathbb{P}(X > 1)$. We see that $\mathbb{P}(X > n) = (1-p)^n$, which is the tail probability of the geometric distribution, and hence $X \sim \text{Geo}(p)$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2020 DIS 13A

1 Continuous Joint Densities

The joint probability density function of two random variables X and Y is given by $f(x,y) = Cxy$ for $0 \leq x \leq 1, 0 \leq y \leq 2$, and 0 otherwise (for a constant C).

(a) Find the constant C that ensures that $f(x,y)$ is indeed a probability density function.

(b) Find $f_X(x)$, the marginal distribution of X.

(c) Find the conditional distribution of Y given $X = x$.

(d) Are X and Y independent?

Solution:

(a) Since $f(x,y)$ is a probability density function, it must integrate to 1. Then:

$$1 = \int_0^1 \int_0^2 Cxy\,dy\,dx = \int_0^1 2Cx dx = C$$

Therefore, $C = 1$.

(b) To get the marginal distribution of X, we integrate the joint distribution with respect to Y. So:

$$f_X(x) = \int_0^2 f(x,y)dy = \int_0^2 xydy = 2x$$

This is the marginal distribution for $0 \leq x \leq 1$.

(c) The conditional distribution of Y given by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{xy}{2x} = \frac{y}{2}$$

(d) The conditional distribution of Y given $X = x$ does not depend on x, so they are independent.

Alternatively, you could find the marginal distribution of Y and see it is the same as the conditional distribution of Y:

$$f_Y(y) = \int_0^1 f(x,y)dx = \int_0^1 xydx = \frac{y}{2}$$

Notice that since X and Y are independent, $f_X(x)f_Y(y) = xy = f_{X,Y}(x,y)$, i.e. the product of the marginal distributions is the same as the joint distribution.

2 Uniform Distribution

You have two fidget spinners, each having a circumference of 10. You mark one point on each spinner as a needle and place each of them at the center of a circle with values in the range $[0, 10)$ marked on the circumference. If you spin both (independently) and let X be the position of the first spinner's mark and Y be the position of the second spinner's mark, what is the probability that $X \geq 5$, given that $Y \geq X$?

Solution:

First we write down what we want and expand out the conditioning:

$$\mathbb{P}[X \geq 5 \mid Y \geq X] = \frac{\mathbb{P}[Y \geq X \cap X \geq 5]}{\mathbb{P}[Y \geq X]}.$$

$\mathbb{P}[Y \geq X] = 1/2$ by symmetry. To find $\mathbb{P}[Y \geq X \cap X \geq 5]$, it helps a lot to just look at the picture of the probability space and use the continuous uniform law $\mathbb{P}[A] = (\text{area of } A)/(\text{area of } \Omega)$. We are interested in the relative area of the region bounded by $x < y < 10$, $5 < x < 10$ to the entire square bounded by $0 < x < 10$, $0 < y < 10$.

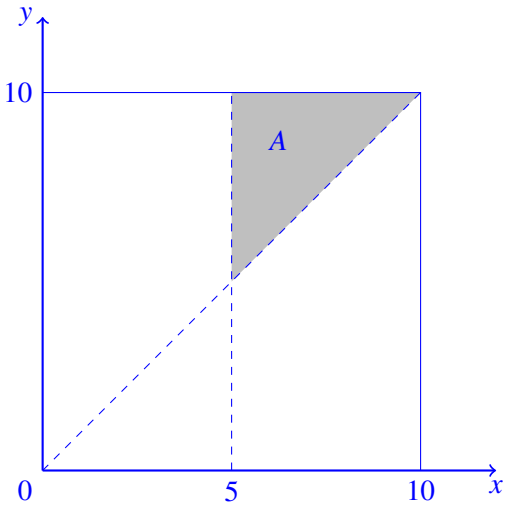


Figure 1: Joint probability density for the spinner.

$$\mathbb{P}[Y \geq X \cap X \geq 5] = \frac{5 \cdot 5/2}{10 \cdot 10} = \frac{1}{8}.$$

So $\mathbb{P}[X \geq 5 \mid Y \geq X] = (1/8)/(1/2) = 1/4$.

3 Darts with Friends

Michelle and Alex are playing darts. Being the better player, Michelle's aim follows a uniform distribution over a circle of radius r around the center. Alex's aim follows a uniform distribution over a circle of radius $2r$ around the center.

(a) Let the distance of Michelle's throw be denoted by the random variable X and let the distance of Alex's throw be denoted by the random variable Y.

- What's the cumulative distribution function of X?
- What's the cumulative distribution function of Y?
- What's the probability density function of X?
- What's the probability density function of Y?

(b) What's the probability that Michelle's throw is closer to the center than Alex's throw? What's the probability that Alex's throw is closer to the center?

(c) What's the cumulative distribution function of $U = \min\{X, Y\}$?

(d) What's the cumulative distribution function of $V = \max\{X, Y\}$?

(e) What is the expectation of the absolute difference between Michelle's and Alex's distances from the center, that is, what is $\mathbb{E}[|X - Y|]$? [*Hint*: Use parts (c) and (d), together with the continuous version of the tail sum formula, which states that $\mathbb{E}[Z] = \int_0^\infty P(Z \geq z)dz$.]

Solution:

(a)

- To get the cumulative distribution function of X, we'll consider the ratio of the area where the distance to the center is less than x, compared to the entire available area. This gives us the following expression:

$$\mathbb{P}(X \leq x) = \frac{\pi x^2}{\pi r^2} = \frac{x^2}{r^2}, \qquad x \in [0, r]$$

- Using the same approach as the previous part:

$$\mathbb{P}(Y \leq y) = \frac{\pi y^2}{\pi \cdot 4r^2} = \frac{y^2}{4r^2}, \qquad y \in [0, 2r]$$

- We'll take the derivative of the CDF to get the following:

$$f_X(x) = \frac{\mathrm{d}\mathbb{P}(X \leq x)}{\mathrm{d}x} = \frac{2x}{r^2}, \qquad x \in [0, r]$$

- Using the same approach as the previous part:

$$f_Y(y) = \frac{\mathrm{d}\mathbb{P}(Y \leq y)}{\mathrm{d}y} = \frac{y}{2r^2}, \qquad y \in [0, 2r]$$

(b) We'll condition on Alex's outcome and then integrate over all the possibilities to get the marginal $\mathbb{P}(X \leq Y)$ as following:

$$\begin{aligned}\mathbb{P}(X \leq Y) &= \int_0^{2r} \mathbb{P}(X \leq Y \mid Y = y) f_Y(y)\,\mathrm{d}y = \int_0^r \frac{y^2}{r^2} \times \frac{y}{2r^2}\,\mathrm{d}y + \int_r^{2r} 1 \times \frac{y}{2r^2}\,\mathrm{d}y \\ &= \frac{r^4 - 0}{8r^4} + \frac{4r^2 - r^2}{4r^2} = \frac{1}{8} + \frac{3}{4} = \frac{7}{8}\end{aligned}$$

Note the range within which $\mathbb{P}(X \leq Y) = 1$. This allowed us to separate the integral to simplify our solution. Using this, we can get $\mathbb{P}(Y \leq X)$ by the following:

$$\mathbb{P}(Y \leq X) = 1 - \mathbb{P}(X \leq Y) = \frac{1}{8}$$

A similar approach to the integral above could be used to verify this result.

$$\mathbb{P}(Y \leq X) = \int_0^r \mathbb{P}(Y \leq X \mid X = x) f_X(x)\,\mathrm{d}x = \int_0^r \frac{x^2}{4r^2}\frac{2x}{r^2}\,\mathrm{d}x = \frac{1}{2r^4}\int_0^r x^3\,\mathrm{d}x = \frac{r^4}{8r^4} = \frac{1}{8}$$

(c) Getting the CDF of U relies on the insight that for the minimum of two random variables to be greater than a value, they both need to be greater than that value. Taking the complement of this will give us the CDF of U. This allows us to get the following result. For $u \in [0, r]$:

$$\begin{aligned}\mathbb{P}(U \leq u) &= 1 - \mathbb{P}(U \geq u) = 1 - \mathbb{P}(X \geq u)\mathbb{P}(Y \geq u) = 1 - \big(1 - \mathbb{P}(X \leq u)\big)\big(1 - \mathbb{P}(Y \leq u)\big) \\ &= 1 - \left(1 - \frac{u^2}{r^2}\right)\left(1 - \frac{u^2}{4r^2}\right) = \frac{5u^2}{4r^2} - \frac{u^4}{4r^4}\end{aligned}$$

For $u > r$, we get $\mathbb{P}(X > u) = 0$, this makes $\mathbb{P}(U \leq u) = 1$.

(d) Getting the CDF of V also relies on a similar insight that for the maximum of two random variables to be smaller than a value, they both need to be smaller than that value. Using this we can get the following result for $v \in [0, r]$:

$$\mathbb{P}(V \leq v) = \mathbb{P}(X \leq v)\mathbb{P}(Y \leq v) = \left(\frac{v^2}{r^2}\right)\left(\frac{v^2}{4r^2}\right) = \frac{v^4}{4r^4}$$

For $v \in [r, 2r]$ we have $\mathbb{P}(X \leq v) = 1$, this makes

$$\mathbb{P}(V \leq v) = \mathbb{P}(Y \leq v) = \frac{v^2}{4r^2}.$$

For $v > 2r$ we have $\mathbb{P}(V \leq v) = 1$ since CDFs of both X and Y are 1 in this range.

(e) We can subtract U from V to get this difference. Using the tail-sum formula to calculate the expectation, we can get the following result:

$$\begin{aligned}\mathbb{E}[|X - Y|] = \mathbb{E}[V - U] = \mathbb{E}[V] - \mathbb{E}[U] &= \int_0^{2r} \mathbb{P}(V \geq v)\,\mathrm{d}v - \int_0^r \mathbb{P}(U \geq u)\,\mathrm{d}u \\ &= \int_0^r \left(1 - \frac{v^4}{4r^4}\right)\mathrm{d}v + \int_r^{2r}\left(1 - \frac{v^2}{4r^2}\right)\mathrm{d}v - \int_0^r \left(1 - \frac{5u^2}{4r^2} + \frac{u^4}{4r^4}\right)\mathrm{d}u \\ &= \frac{19r}{20} + \frac{5r}{12} - \frac{19r}{30} = \frac{11r}{15}\end{aligned}$$

Alternatively, you could derive the density of U and V and use those to calculate the expectation. For $v \in [0, r]$:

$$f_V(v) = \frac{\mathrm{d}\mathbb{P}(V \leq v)}{\mathrm{d}v} = \frac{v^3}{r^4}$$

For $v \in [r, 2r]$:

$$f_V(v) = \frac{\mathrm{d}\mathbb{P}(V \leq v)}{\mathrm{d}v} = \frac{v}{2r^2}$$

Using this we can calculate $\mathbb{E}[V]$ as:

$$\mathbb{E}[V] = \int_0^{2r} v f_V(v)\,\mathrm{d}v = \frac{1}{r^4}\int_0^r v^4\,\mathrm{d}v + \frac{1}{2r^2}\int_r^{2r} v^2\,\mathrm{d}v = \frac{r^5}{5r^4} + \frac{8r^3 - r^3}{6r^2} = \frac{r}{5} + \frac{7r}{6r} = \frac{41r}{30}$$

To calculate $\mathbb{E}[U]$ we will use the following PDF for $u \in [0, r]$:

$$f_U(u) = \frac{\mathrm{d}\mathbb{P}(U \leq u)}{\mathrm{d}u} = \frac{5u}{2r^2} - \frac{u^3}{r^4}$$

We can get the $\mathbb{E}[U]$ by the following:

$$\mathbb{E}[U] = \int_0^r u f_U(u)\,\mathrm{d}u = \int_0^r \left(\frac{5u^2}{2r^2} - \frac{u^4}{r^4}\right)\mathrm{d}u = \frac{5r^3}{6r^2} - \frac{r^5}{5r^4} = \frac{5r}{6} - \frac{r}{5} = \frac{19r}{30}$$

Combining the two results gives us the same result as above:

$$\mathbb{E}[|X - Y|] = \mathbb{E}[V - U] = \mathbb{E}[V] - \mathbb{E}[U] = \frac{41r}{30} - \frac{19r}{30} = \frac{11r}{15}$$

1 First Exponential to Die

Let X and Y be Exponential(λ_1) and Exponential(λ_2) respectively, independent. What is

$$\mathbb{P}\big(\min(X,Y) = X\big),$$

the probability that the first of the two to die is X?

Solution:

Recall that the CDF of an exponential is $\mathbb{P}[X \le x] = 1 - \exp(-\lambda x)$ for $x \ge 0$.

$$\mathbb{P}\big(\min(X,Y) = X\big) = \mathbb{P}(Y > X) = \int_0^\infty \mathbb{P}(Y > X \mid X = x) f_X(x)\,dx = \int_0^\infty e^{-\lambda_2 x} \cdot \lambda_1 e^{-\lambda_1 x}\,dx$$
$$= -\frac{\lambda_1}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)x}\Big|_{x=0}^{\infty} = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

2 Chebyshev's Inequality vs. Central Limit Theorem

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \ge 2]$.

(c) Can you use b to bound $\mathbb{P}[Z_n \ge 2]$ and $\mathbb{P}[Z_n \le -2]$?

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \le 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, can you provide approximations for $\mathbb{P}[Z_n \ge 2]$ and $\mathbb{P}[Z_n \le -2]$?

Solution:

(a) $\mathbb{E}[X_1] = -1/12 + 9/12 + 4/12 = 1$, and

$$\operatorname{Var} X_1 = \frac{1}{12} \cdot 2^2 + \frac{9}{12} \cdot 0^2 + \frac{2}{12} \cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \ldots, X_n$ are independent), we find that $\mathbb{E}[\sum_{i=1}^n X_i] = n$ and $\operatorname{var}(\sum_{i=1}^n X_i) = n/2$.

Again, by linearity of expectation, $\mathbb{E}[\sum_{i=1}^n X_i - n] = n - n = 0$. Subtracting a constant does not change the variance, so $\operatorname{var}(\sum_{i=1}^n X_i - n) = n/2$, as before.

Using the scaling properties of the expectation and variance, $\mathbb{E}[Z_n] = 0/\sqrt{n/2} = 0$ and $\operatorname{Var} Z_n = (n/2)/(n/2) = 1$.

(b)

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\operatorname{Var} Z_n}{2^2} = \frac{1}{4}$$

(c) $1/4$ for both, since $\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$ and $\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$.

(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.

3 Why Is It Gaussian?

Let X be a normally distributed random variable with mean μ and variance σ^2. Let $Y = aX + b$, where $a > 0$ and b are non-zero real numbers. Show explicitly that Y is normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$. The PDF for the Gaussian Distribution is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. One approach is to start with the cumulative distribution function of Y and use it to derive the probability density function of Y.

[1.You can use without proof that the pdf for any gaussian with mean and sd is given by the formula $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ where μ is the mean value for X and σ^2 is the variance. 2. The drivative of CDF gives PDF.]

Solution:

Problem and solution taken from *A First Course in Probability* by Sheldon Ross, 8th edition.

Let $a > 0$.

We start with the cumulative distribution function (CDF) of Y, F_Y.

$$\begin{aligned} F_Y(x) &= \mathbb{P}[Y \le x] && \text{By definition of CDF} \\ &= \mathbb{P}[aX + b \le x] && \text{Plug in } Y = aX + b \\ &= \mathbb{P}\left[X \le \frac{x-b}{a}\right] && \text{Because } a > 0 \\ &= F_X\left(\frac{x-b}{a}\right) && \text{By definition of CDF. } F_X \text{ denotes the CDF of } X. \end{aligned} \tag{1}$$

Let f_Y denote the probability density function (PDF) of Y.

$$\begin{aligned} f_Y(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_Y(x) && \text{The PDF is the derivative of the CDF.} \\ &= \frac{\mathrm{d}}{\mathrm{d}x} F_X\left(\frac{x-b}{a}\right) && \text{Plug in the result from (1)} \\ &= \frac{1}{a} \cdot f_X\left(\frac{x-b}{a}\right) && \text{PDF is the derivative of CDF.} \\ & && \text{Apply chain rule, } \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x-b}{a}\right) = \frac{1}{a}. \\ &= \frac{1}{a} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-((x-b)/a-\mu)^2/(2\sigma^2)} && X \sim \mathcal{N}(\mu, \sigma^2). \\ &= \frac{1}{a\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-(x-b-a\mu)^2/(2\sigma^2 a^2)} && \frac{x-b}{a} - \mu = \frac{1}{a}(x-b-a\mu) \end{aligned} \tag{2}$$

We have shown that f_Y equals the probability density function of a normal random variable with mean $b + a\mu$ and variance $\sigma^2 a^2$. So, Y is normally distributed with mean $b + a\mu$ and variance $\sigma^2 a^2$. The proof is done for $a > 0$. The proof for $a < 0$ is similar.

1 Markov Chains: Prove/Disprove

Prove or disprove the following statements, using the definitions from the previous question.

(a) There exists an irreducible, finite Markov chain for which there exist initial distributions that converge to different distributions.

(b) There exists an irreducible, aperiodic, finite Markov chain for which $\mathbb{P}(X_{n+1} = j \mid X_n = i) = 1$ or 0 for all i, j.

(c) There exists an irreducible, non-aperiodic Markov chain for which $\mathbb{P}(X_{n+1} = j \mid X_n = i) \neq 1$ for all i, j.

(d) For an irreducible, non-aperiodic Markov chain, any initial distribution not equal to the invariant distribution does not converge to any distribution.

Solution:

(a) False. Every finite irreducible Markov chain has a unique stationary distribution. If it's possible for the Markov chain to converge to two different distributions given different starting distributions, it implies there are two stationary distributions. To elaborate further, we know in the long run the fraction of time spent in each state converges to the stationary distribution. So if the distribution converges, the long-run fraction of time will be whatever distribution it converges to, which we see must be the stationary distribution.

(b) True, you can have one state pointing to itself. However for number of states > 1 it is false. Consider the initial distribution of having a probability of 1 of being in an arbitrary state. After a transition, the resulting distribution must be a probability 1 of being in a different state (if it were the same state, this would immediately imply that the Markov chain is reducible). Further transitions have the same effect. Therefore this initial distribution does not converge. Therefore this Markov chain cannot be aperiodic and irreducible (since it would converge in that case).

(c) True. Consider the states $\{0,1,2,3\}$. Set $P(i, j) = 1/2$ if $i \equiv j \pm 1 \mod 4$ and 0 otherwise. In other words, the Markov chain is a square with each side replaced with two links pointing in opposite directions with probabilities of $1/2$. Consider the period of state 0. Any path from 0 back to itself, such as $0-1-2-1-0$, alternates in parity of each consecutive state since each state only points to the state above or below it mod 4. Therefore state 0 has period 2. Therefore this Markov chain is not aperiodic (and all states have period 2).

(d) False. Take the initial distribution $\begin{bmatrix}0.25 & 0.30 & 0.25 & 0.20\end{bmatrix}$ for the above Markov chain. After one transition it goes to the invariant distribution, $\begin{bmatrix}0.25 & 0.25 & 0.25 & 0.25\end{bmatrix}$.

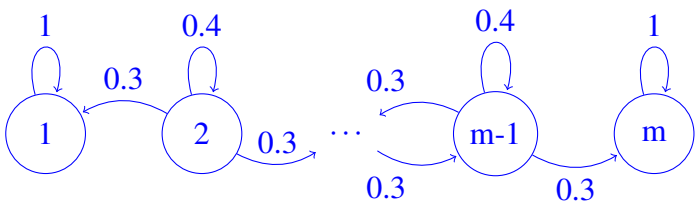
2 Can it be a Markov Chain?

(a) A fly flies in a straight line in unit-length increments. Each second it moves to the left with probability 0.3, right with probability 0.3, and stays put with probability 0.4. There are two spiders at positions 1 and m and if the fly lands in either of those positions it is captured. Given that hte fly starts between positions 1 and m, model this process as a Markov Chain.

(b) Take the same scenario as in the previous part with $m = 4$. Let $Y_n = 0$ if at time n the fly is in position 1 or 2 and let $Y_n = 1$ if at time n the fly is in position 3 or 4. Is the process Y_n a Markov chain?

Solution:

(a) We can draw the Markov chain as such:



(b) No, because the longer the fly stays in any one state, the more likely the fly gets in one of the absorbing states.

For example, say $\mathbb{P}[X_0 = 2] = \mathbb{P}[X_0 = 3] = 1/2$ and $\mathbb{P}[X_0 = 1] = \mathbb{P}[X_0 = 4] = 0$. Then

$$\begin{aligned}
\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 0] &= \mathbb{P}[X_2 \in \{1,2\} \mid X_1 = 3, X_0 = 2] \\
&= \mathbb{P}[X_2 = 2 \mid X_1 = 3] = 0.3 \\
\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 1] &= \mathbb{P}[Y_2 = 0, Y_1 = 1, Y_0 = 1] / \mathbb{P}[Y_1 = 1, Y_0 = 1] \\
&= \mathbb{P}[X_2 = 2, X_1 = 3, X_0 = 3] / (\mathbb{P}[X_1 = 3, X_0 = 3] + \mathbb{P}[X_1 = 4, X_0 = 3]) \\
&= \frac{0.5 \cdot 0.4 \cdot 0.3}{0.5 \cdot 0.4 + 0.5 \cdot 0.3} = \frac{6}{35}
\end{aligned}$$

If Y was Markov, then $\mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 0] = \mathbb{P}[Y_2 = 0 \mid Y_1 = 1] = \mathbb{P}[Y_2 = 0 \mid Y_1 = 1, Y_0 = 1]$. However, $0.3 > 6/35$, and so Y cannot be Markov.

3 Allen's Umbrella Setup

Every morning, Allen walks from his home to Soda, and every evening, Allen walks from Soda to his home. Suppose that Allen has two umbrellas in his possession, but he sometimes leaves his

umbrellas behind. Specifically, before leaving from his home or Soda, he checks the weather. If it is raining outside, he will bring his umbrella (that is, if there is an umbrella where he currently is). If it is not raining outside, he will forget to bring his umbrella. Assume that the probability of rain is p.

(a) Model this as a Markov chain. What is $\mathscr{X}$? Write down the transition matrix.

(b) What is the transition matrix after 2 trips? n trips? Determine if the distribution of X_n converges to the invariant distribution, and compute the invariant distribution. Determine the long-term fraction of time that Allen will walk through rain with no umbrella.

Solution:

(a) Suppose Allen is in state 0. Then, Allen has no umbrellas to bring, so with probability 1 Allen arrives at a location with 2 umbrellas. That is,

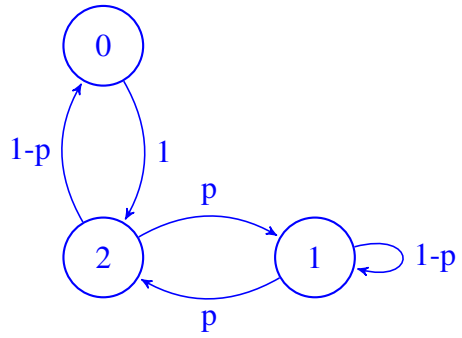
$$\mathbb{P}[X_{n+1} = 2 \mid X_n = 0] = 1.$$

Suppose Allen is in state 1. With probability p, it rains and Allen brings the umbrella, arriving at state 2. With probability $1-p$, Allen forgets the umbrella, so Allen arrives at state 1.

$$\mathbb{P}[X_{n+1} = 2 \mid X_n = 1] = p, \qquad \mathbb{P}[X_{n+1} = 1 \mid X_n = 1] = 1-p$$

Suppose Allen is in state 2. With probability p, it rains and Allen brings the umbrella, arriving at state 1. With probability $1-p$, Allen forgets the umbrella, so Allen arrives at state 0.

$$\mathbb{P}[X_{n+1} = 1 \mid X_n = 2] = p, \qquad \mathbb{P}[X_{n+1} = 0 \mid X_n = 2] = 1-p$$



We summarize this with the transition matrix

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix}.$$

(b) The transition matrices would be expressed as P^2 and P^n. Below we find the stationary distribution.

Observe that the transition matrix has non-zero element in its diagonal, which means the minimum number of steps to transit to state 1 from itself is one. Thus this transition matrix is irreducible and aperiodic, so it converges to its invariant distribution. To solve for the distribution, we set $\pi P = \pi$, or $\pi(P - I) = 0$. This yields the balance equations

$$\begin{bmatrix}\pi(0) & \pi(1) & \pi(2)\end{bmatrix}\begin{bmatrix}-1 & 0 & 1\\ 0 & -p & p\\ 1-p & p & -1\end{bmatrix} = \begin{bmatrix}0 & 0 & 0\end{bmatrix}.$$

As usual, one of the equations is redundant. We replace the last column by the normalization condition $\pi(0)+\pi(1)+\pi(2) = 1$.

$$\begin{bmatrix}\pi(0) & \pi(1) & \pi(2)\end{bmatrix}\begin{bmatrix}-1 & 0 & 1\\ 0 & -p & 1\\ 1-p & p & 1\end{bmatrix} = \begin{bmatrix}0 & 0 & 1\end{bmatrix}$$

Now solve for the distribution:

$$\begin{bmatrix}\pi(0) & \pi(1) & \pi(2)\end{bmatrix} = \frac{1}{3-p}\begin{bmatrix}1-p & 1 & 1\end{bmatrix}$$

The invariant distribution also tells us the long-term fraction of time that Allen spends in each state. We can see that Allen spends a fraction $(1-p)/(3-p)$ of his time with no umbrella in his location, so the long-term fraction of time in which he walks through rain is $p(1-p)/(3-p)$.

4 Three Tails

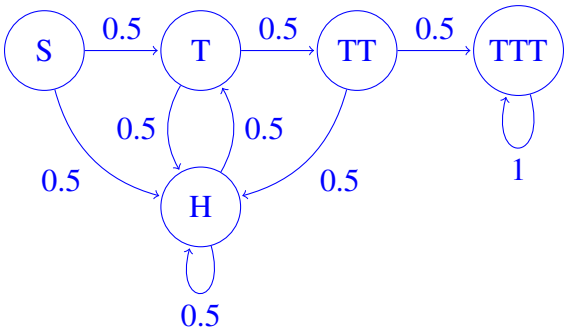
You flip a fair coin until you see three tails in a row. What is the average number of heads that you'll see until getting TTT?

Hint: How is this different than the number of *coins* flipped until getting TTT?

Solution:

We can model this problem as a Markov chain with the following states:

- S: Start state, which we are only in before flipping any coins.
- H: We see a head, which means no streak of tails currently exists.
- T: We've seen exactly one tail in a row so far.
- TT: We've seen exactly two tails in a row so far.
- TTT: We've accomplished our goal of seeing three tails in a row and stop flipping.



We can write the first step equations and solve for $\beta(S)$, only counting heads that we see since we are not looking for the total number of flips. The equations are as follows:

$$\beta(S) = 0.5\beta(T) + 0.5\beta(H) \tag{1}$$
$$\beta(H) = 1 + 0.5\beta(H) + 0.5\beta(T) \tag{2}$$
$$\beta(T) = 0.5\beta(TT) + 0.5\beta(H) \tag{3}$$
$$\beta(TT) = 0.5\beta(H) + 0.5\beta(TTT) \tag{4}$$
$$\beta(TTT) = 0 \tag{5}$$

From equation (2), we see that

$$0.5\beta(H) = 1 + 0.5\beta(T)$$

and can substitute that into equation (3) to get

$$0.5\beta(T) = 0.5\beta(TT) + 1.$$

Substituting this into equation (4), we can deduce that $\beta(TT) = 4$. This allows us to conclude that $\beta(T) = 6$, $\beta(H) = 8$, and $\beta(S) = 7$. On average, we expect to see 7 heads before flipping three tails in a row.