CS 70 Spring 2024 | Discrete Mathematics and Probability Theory
Seshia, Sinclair | DIS 0A

CS 70 Spring 2024 — Discrete Mathematics and Probability Theory — Seshia, Sinclair — DIS 0A

1 Propositional Practice

Note 1

Convert the following English sentences into propositional logic and the following propositions into English. State whether or not each statement is true with brief justification.

(a) There is a real number which is not rational.

(b) All integers are natural numbers or are negative, but not both.

(c) If a natural number is divisible by 6, it is divisible by 2 or it is divisible by 3.

(d) $(\forall x \in \mathbb{Z})\ (x \in \mathbb{Q})$

(e) $(\forall x \in \mathbb{Z})\ (((2 \mid x) \vee (3 \mid x)) \implies (6 \mid x))$

(f) $(\forall x \in \mathbb{N})\ ((x > 7) \implies ((\exists a, b \in \mathbb{N})\ (a + b = x)))$

Solution:

(a) $(\exists x \in \mathbb{R})\ (x \notin \mathbb{Q})$, or equivalently $(\exists x \in \mathbb{R})\ \neg(x \in \mathbb{Q})$. This is true, and we can use π as an example to prove it.

(b) $(\forall x \in \mathbb{Z})\ (((x \in \mathbb{N}) \vee (x < 0)) \wedge \neg((x \in \mathbb{N}) \wedge (x < 0)))$. This is true, since we define the naturals to contain all integers which are not negative.

(c) $(\forall x \in \mathbb{N})\ ((6 \mid x) \implies ((2 \mid x) \vee (3 \mid x)))$. This is true, since any number divisible by 6 can be written as $6k = (2 \cdot 3)k = 2(3k)$, meaning it must also be divisible by 2.

(d) All integers are rational numbers. This is true, since any integer number n can be written as $n/1$.

(e) Any integer that is divisible by 2 or 3 is also divisible by 6. This is false–2 provides the easiest counterexample. Note that this statement is false even though its converse (part c) is true.

(f) If a natural number is larger than 7, it can be written as the sum of two other natural numbers. This is trivially true, since we can take $a = x$ and $b = 0$.

(Aside: this is a refererence to the very weak Goldback Conjecture (https://xkcd.com/1310/).)

2 Truth Tables

Note 1

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

3 Implication

Note 0
Note 1

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.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 0B

1 Perfect Square

Note 2

(a) Prove that if n^2 is odd, then n must also be odd.

(b) Prove that if n^2 is odd, then n^2 can be written in the form $8k+1$ for some integer k.

Solution:

(a) We will proceed by a proof by contraposition; the contrapositive of the statement is "if n is even, then n^2 is also even". Here, since n is even, we can write $n = 2k$ for some integer k. This makes $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$, which is even, as desired. By contraposition, this means that if n^2 is odd, then n must also be odd.

(b) We will proceed with a direct proof. From the previous part, since n^2 is odd, n is also odd, i.e., of the form $n = 2l + 1$ for some integer l. Then, $n^2 = 4l^2 + 4l + 1 = 4l(l+1) + 1$. Since one of l and $l+1$ must be even, $l(l+1)$ is of the form $2k$ for some integer k and $n^2 = 8k+1$.

2 Numbers of Friends

Note 2

Prove that if there are $n \geq 2$ people at a party, then at least 2 of them have the same number of friends at the party. Assume that friendships are always reciprocated: that is, if Alice is friends with Bob, then Bob is also friends with Alice.

(Hint: The Pigeonhole Principle states that if n items are placed in m containers, where $n > m$, at least one container must contain more than one item. You may use this without proof.)

Solution:

We will prove this by contradiction. Suppose the contrary that everyone has a different number of friends at the party. Since the number of friends that each person can have ranges from 0 to $n-1$, we conclude that for every $i \in \{0, 1, \dots, n-1\}$, there is exactly one person who has exactly i friends at the party. In particular, there is one person who has $n-1$ friends (i.e., friends with everyone), is friends with a person who has 0 friends (i.e., friends with no one). This is a contradiction since friendship is mutual.

Here, we used the pigeonhole principle because assuming for contradiction that everyone has a different number of friends gives rise to n possible containers. Each container denotes the number of friends that a person has, so the containers can be labelled $0, 1, ..., n-1$. The objects assigned to these containers are the people at the party. However, containers 0, $n-1$ or both must be empty since these two containers cannot be occupied at the same time. This means that we are assigning n people to at most $n-1$ containers, and by the pigeonhole principle, at least one of the $n-1$ containers has to have two or more objects i.e. at least two people have to have the same number of friends.

3 Pebbles

Note 2

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; the contrapositive is "if there does not exist an all-red column, then there is always a way of choosing one pebble from each column such that there does not exist a red pebble among the chosen ones".

Suppose there does not exist an all-red column. This means that we can always find a blue pebble in each column. 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.

We can also approach the problem through contradiction; the logic stays almost exactly the same, and we start with the negation of the conclusion: that there does not exist an all-red column. The same reasoning above allows us to conclude that there will always exist a way of choosing one pebble from each column such that all pebbles are blue (i.e. no pebbles are red).

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair **DIS 1A**

1 Natural Induction on Inequality

Note 3

Prove that if $n \in \mathbb{N}$ and $x > 0$, then $(1+x)^n \geq 1+nx$.

Solution:

- *Base Case:* When $n = 0$, the claim holds since $(1+x)^0 \geq 1+0x$.
- *Inductive Hypothesis:* Assume that $(1+x)^k \geq 1+kx$ for some value of $n = k$ where $k \in \mathbb{N}$.
- *Inductive Step:* For $n = k+1$, we can show the following:

$$\begin{aligned}(1+x)^{k+1} = (1+x)^k(1+x) &\geq (1+kx)(1+x)\\ &\geq 1+kx+x+kx^2\\ &\geq 1+(k+1)x+kx^2 \geq 1+(k+1)x\end{aligned}$$

By induction, we have shown that $\forall n \in \mathbb{N}, (1+x)^n \geq 1+nx$.

2 Make It Stronger

Note 3

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 inductive hypothesis be simply $a_n \leq 3^{(2^n)}$? Attempt an induction proof with this hypothesis to show why this does not work.

(b) Try to instead prove the statement $a_n \leq 3^{(2^n-1)}$ using induction.

(c) Why does the hypothesis in part (b) imply the overall claim?

Solution:

(a) Let's 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 $n+1$.

(c) 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 in part (b) does indeed imply what we wanted to prove in part (a).

3 Binary Numbers

Note 3

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

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

for some $k \in \mathbb{N}$ and $c_i \in \{0, 1\}$ for all $i \leq k$.

Solution:

Prove by strong induction on n.

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.

Here is another alternate solution emulating the algorithm of converting a decimal number to a binary number.

- Base Case: $n = 1$ can be written as 1×2^0.
- Inductive Step: Assume that the statement is true for all $1 \leq m \leq n$, for arbitrary n. We show that the statement holds for $n+1$. Let 2^m be the largest power of 2 such that $n+1 \geq 2^m$. Thus, $n+1 < 2^{m+1}$. We examine the number $(n+1)-2^m$. Since $(n+1)-2^m < n+1$, the inductive hypothesis holds, so we have a binary representation for $(n+1)-2^m$. (If $(n+1)-2^m = 0$, then we still have a binary representation, namely $0 \cdot 2^0$.)

  Also, since $n+1 < 2^{m+1}$, $(n+1)-2^m < 2^m$, so the largest power of 2 in the representation of $(n+1)-2^m$ is 2^{m-1}. Thus, by the inductive hypothesis,

  $$(n+1)-2^m = c_{m-1} \cdot 2^{m-1} + c_{m-2} \cdot 2^{m-2} + \cdots + c_1 \cdot 2^1 + c_0 \cdot 2^0,$$

  and adding 2^m to both sides gives

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

  which is a binary representation for $n+1$. Thus, the induction is complete.

Another intuition is that if x has a binary representation, $2x$ and $2x+1$ do as well: shift the bits and possibly place 1 in the last bit. The above induction could then have proceeded from n and used the binary representation of $\lfloor n/2 \rfloor$, shifting and possibly setting the first bit depending on whether n is odd or even.

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 $P(n+1)$.

4 Fibonacci for Home

Note 3

Recall, the Fibonacci numbers, defined recursively as

$F_1 = 1$, $F_2 = 1$, and $F_n = F_{n-2} + F_{n-1}$.

Prove that every third Fibonacci number is even. For example, $F_3 = 2$ is even and $F_6 = 8$ is even.

Solution:

We want to prove that for all natural numbers $k \geq 1$, F_{3k} is even.

Base case: For $k = 1$, we can see that $F_3 = 2$ is even.

Induction hypothesis: Suppose that for an arbitrary fixed value of k, F_{3k} is even.

Inductive step: We can write

$$F_{3k+3} = F_{3k+2} + F_{3k+1} = 2F_{3k+1} + F_{3k}.$$

By the induction hypothesis, we know that $F_{3k} = 2q$ for some q.

This means that we have that $F_{3k+3} = 2(F_{3k+1} + q)$, which implies that it is even. Thus, by the principles of induction we have shown that all F_{3k} are even.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 1B

1 Stable Matching

Note 4

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

Jobs	Candidates
1	A > B > C
2	B > A > C
3	A > B > C

Candidates	Jobs
A	2 > 1 > 3
B	1 > 3 > 2
C	1 > 2 > 3

Run the traditional 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 5 days to produce a matching. The resulting pairing is as follows. The circles indicate the job that a candidate picked on a given day (and rejected the rest).

$$\{(A,2),(B,1),(C,3)\}.$$

Candidate	Day 1	Day 2	Day 3	Day 4	Day 5
A	①,3	①	1,②	②	②
B	②	2,③	③	①,3	①
C					③

2 Propose-and-Reject Proofs

Note 4

Prove the following statements about the traditional propose-and-reject algorithm.

(a) In any execution of the algorithm, if a candidate receives a proposal on day i, then she receives some proposal on every day thereafter until termination.

(b) In any execution of the algorithm, if a candidate receives no proposal on day i, then she receives no proposal on any previous day j, $1 \leq j < i$.

(c) In any execution of the algorithm, there is at least one candidate who only receives a single proposal. (Hint: use the parts above!)

Solution:

(a) The idea is to induct on the number of days passed so far.
Base case: Candidate C receives a proposal on day i
Inductive Step: Assume C receives a proposal on day $j \geq i$ from job J. We want to show she will also get a proposal on day $j+1$. There are two cases: C prefers J to all other offers, or C prefers some job J' to J. In the first case J proposes to C on day $j+1$ and in the second J' proposes to C on day $j+1$ so C receives at least one proposal on day $j+1$.

(b) One way is to use a proof by contradiction. Assume that a candidate receives no proposal on day i but did receive a proposal on some previous day j, $1 \leq j < i$. By the previous part, since the candidate received a proposal on day j, she must receive at least one proposal on every day after j. But $i > j$, so the candidate must have received a proposal on day i, contradicting our original assumption that she did not.

(c) Let's say the algorithm takes k days. This means that every candidate must have received a proposal on day k. However, this also means that there is at least one candidate C who does not receive a proposal on day $k-1$–if this were not the case, the algorithm would have already terminated on day $k-1$. Then from part (b), since C did not receive a proposal on day $k-1$, she didn't receive a proposal on any day before k. Furthermore, we know she got exactly one proposal on day k, since the algorithm terminated on that day. Thus, we have that C receives exactly one proposal throughout the entire run of the algorithm.

3 Be a Judge

Note 4

By stable matching instance, we mean a set of jobs and candidates and their preference lists. For each of the following statements, indicate whether the statement is True or False and justify your answer with a short 2-3 line explanation:

(a) There is a stable matching instance for n jobs and n candidates for $n > 1$, such that in a stable matching algorithm with jobs proposing, every job ends up with its least preferred candidate.

(b) In a stable matching instance, if job J and candidate C each put each other at the top of their respective preference lists, then J must be paired with C in every stable pairing.

(c) In a stable matching instance with at least two jobs and two candidates, if job J and candidate C each put each other at the bottom of their respective preference lists, then J cannot be paired with C in any stable pairing.

(d) For every $n > 1$, there is a stable matching instance for n jobs and n candidates which has an **unstable** pairing where **every** unmatched job-candidate pair is a rogue couple or pairing.

Solution:

(a) **False**: If this were to occur, it would mean that at the end of the algorithm, every job would have proposed to every candidate on its list and has been rejected $n-1$ times. This also means that every candidate got proposed to by every single one of the jobs, and at the very end must have rejected $n-1$ jobs in total (to end up with only one preferred job at the last day). We know this is impossible though, as we learned above that at least one candidate receives a single proposal. Thus, there must be at least one candidate who is not proposed to until the very last day.

(b) **True**: We give a simple proof by contradiction. Assume that J and C put each other at the top of their respective preference lists, but J and C are not paired with each other in some stable pairing S. Thus, S includes the pairings (J, C'), (J', C), for some job J' and candidate C'. However, J prefers C over its partner in S, since C is at the top of J's preference list. Similarly C prefers J over her current job. Thus (J,C) form a rogue couple in S, so S is not stable. We have arrived at a contradiction that S exists where C and J are not paired.

Therefore if job J and candidate C put each other at the top of their respective preference lists, then J must be paired with C in every stable pairing.

(c) **False**: The key here is to realize that this is possible if job J and candidate C are at the bottom of everybody else's preference list as well. For example, a two job, two candidate instance where the first job is best for both candidates, and the first candidate is best for both jobs has its only stable pairing where the first job and first candidate are paired (by part (b)), and the second job and second candidate are paired. The second job and second candidate are each other's least preferred option.

(d) **True**: The key idea to this solution is that we want a set of preferences for which J_i and C_i like each other the least and make a pairing J_i and C_i together. In this matching M each unmatched couple is a rogue couple. In particular, an instance can be constructed by forming preferences where job i's (candidate i's) least favorite candidate is candidate i (job i and an *arbitrary* ordering on all the others. Thus, for any job i and candidate j, where $i \neq j$, then job i prefers candidate j to i (its partner in M) and candidate j prefers job i to j (its partner in M.) That, is every pair i and j is a rogue couple.

An example for two jobs and two candidates, is the instance:

Jobs		
1	A	B
2	B	A

Candidates		
A	1	2
B	2	1

The pairing $P = \{(A,2),(B,1)\}$ is the pairing where each pair of jobs and candidates that are not in P, $(A,1)$ and $(B,2)$, are rogue couples.

4 Stable Matching III

Note 4

(a) True or False?

(i) If a candidate accidentally rejects a job she prefers on a given day, then the algorithm still always ends with a matching.

(ii) The Propose-and-Reject Algorithm never produces a candidate-optimal matching.

(iii) If the same job is last on the preference list of every candidate, the job must end up with its least preferred candidate.

(b) As you've seen from lecture, the jobs-proposing Propose-and-Reject Algorithm produces an employer-optimal stable matching. Let's see if the candidate have any way of improving their standing. Suppose exactly one of the candidates has the power to arbitrarily reject one proposal, regardless of which job she has on her string (if any). Construct an example that illustrates the following: for any $n \geq 2$, there exists a stable matching instance for which using this power helps **every** candidate, i.e. every candidate gets a better job than she would have gotten under the jobs-proposing Propose-and-Reject Algorithm.

Solution:

(a) (i) False, consider the case:

Jobs		
1	A	B
2	B	A

Candidates		
A	1	2
B	2	1

Using SMA, the matching will be: $(A,1),(B,2)$. If candidate A rejects job 1 despite having no other jobs on her string, job 1 will propose to candidate B and also get rejected. This leaves both candidate A and job 1 partnerless. In this case, the accidental rejection prevents a matching from being produced at all.

(ii) False. Suppose that all jobs have a different first choice. Also supposed that the job proposing is each candidate's first choice. In this case, the algorithm would end after the first day with both jobs and candidates ending with their top pick. In this case, the result is candidate-optimal.

(iii) False, consider the following case where jobs are numbers and candidates are letters:

Jobs		
1	A	B
2	B	A

Candidates		
A	2	1
B	2	1

Job 1 is last on every candidate's list, however, $\{(A,1),(B,2)\}$ is a stable pairing where job 1 got its top choice candidate.

(b) Without loss of generality, assume that candidate 1 is the candidate with this special power. Now, assume the preference lists are ordered as follows:

Job	Preferences	Candidate	Preferences
1	$1 > 2 > \cdots > n-1 > n$	1	$n > n-1 > \cdots > 2 > 1$
2	$2 > 3 > \cdots > n-1 > n > 1$	2	$1 > n > n-1 > \cdots > 2$
3	$3 > 4 > \cdots > n > 1 > 2$	3	$2 > 1 > n > n-1 > \cdots > 3$
	$\cdots$		$\cdots$
n	$n > 1 > 2 > \cdots > n-1$	n	$n-1 > n-2 > \cdots > 1 > n$

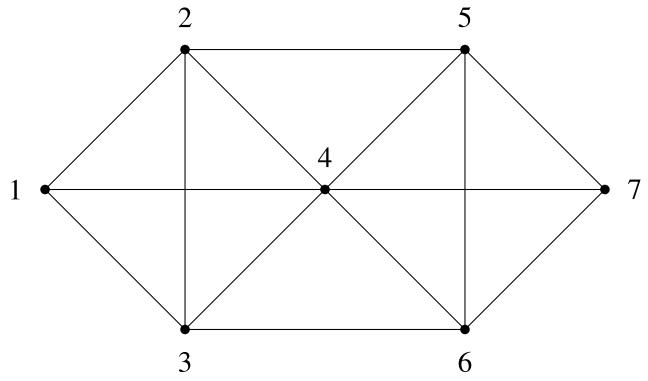
If the Propose-and-Reject Algorithm was run with these preference lists, then each candidate would be stuck with her least-preferred job. However, let's say candidate 1 rejects job 1 on the first day, even though she has nobody on her string. Then, job 1 will be forced to propose to its second option, candidate 2, and she will accept because the job is her first choice. Now, job 2 has no partner and will propose to candidate 3, who will accept, leaving job 3 without a partner. This process continues until job n proposes to candidate 1. One rejection has led all candidates to get their best choice instead of their worst choice!

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 2A

1 Eulerian Tour and Eulerian Walk

Note 5



(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: $1 \to 2 \to 3 \to 4 \to 1 \to 3 \to 6 \to 4 \to 5 \to 2 \to 4 \to 7 \to 5 \to 6 \to 7$ 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). 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 Coloring Trees

Note 5

(a) Prove that all trees with at least 2 vertices have at least two leaves. Recall that a leaf is defined as a node in a tree with degree exactly 1.

(b) Prove that all trees with at least 2 vertices are *bipartite*: the vertices can be partitioned into two groups so that every edge goes between the two groups.

[*Hint:* Use induction on the number of vertices.]

Solution:

(a) For an arbitrary tree $T=(V,E)$ where $|V|\geq 2$, suppose L denotes the set of leaves. This means we have $|V|-|L|$ non leaves. A leaf has degree 1 and the other vertices must have degree at least 2. Moreover, we know that an $|V|$-vertex tree must have $|V|-1$ edges. By the Handshake Lemma,

$$2|V|-2=\sum_{v\in V}\deg(v)=\sum_{v\in L}\deg(v)+\sum_{v\in V\setminus L}\deg(v)\geq |L|+2\left(|V|-|L|\right)=2|V|-|L|$$

which implies that $|L|\geq 2$ as desired.

(b) Proof using induction on the number of vertices n.

Base case $n=2$. A tree with two vertices has only one edge and is a bipartite graph by partitioning the two vertices into two separate parts.

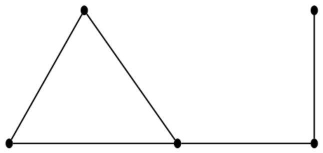
Inductive hypothesis. Assume that all trees with k vertices for an arbitrary $k\geq 2$ is bipartite.

Inductive step. Consider a tree $T=(V,E)$ with $k+1$ vertices. We know that every tree must have at least two leaves (previous part), so remove one leaf u and the edge connected to u, say edge e. The resulting graph $T-u$ is a tree with k vertices and is bipartite by the inductive hypothesis. Thus there exists a partitioning of the vertices $V=R\cup L$ such that there does not exist an edge that connects two vertices in L or two vertices in R. Now when we add u back to the graph. If edge e connects u with a vertex in L then let $L'=L$ and $R'=R\cup\{u\}$. On the other hand if edge e connects u with a vertex in R then let $L'=L\cup\{u\}$ and $R'=R$. L' and R' gives us the required partition to show that T is bipartite. This completes the inductive step and hence by induction we get that all trees with at least 2 vertices are bipartite.

3 Degree Sequences

Note 5

The *degree sequence* of a graph is the sequence of the degrees of the vertices, arranged in descending order, with repetitions as needed. For example, the degree sequence of the following graph is $(3,2,2,2,1)$.



For each of the parts below, determine if there exists a simple undirected graph G (i.e. a graph without self-loops and multiple-edges) having the given degree sequence. Justify your claim.

(a) $(3,3,2,2)$

(b) $(3,2,2,2,2,1,1)$

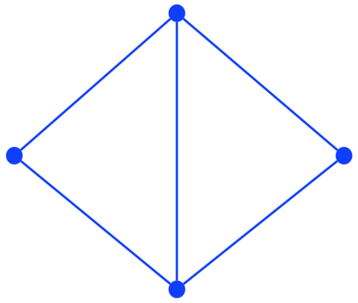
(c) $(6,2,2,2)$

(d) $(4,4,3,2,1)$

Solution:

(a) **Yes**

The following graph has degree sequence $(3,3,2,2)$.



(b) **No**

For any graph G, the number of vertices that have odd degree is even (since the sum of degrees is twice the number of edges). The given degree sequence has 3 odd degree vertices.

(c) **No**

The total number of vertices is 4. Hence there cannot be a vertex with degree 6.

(d) **No**

The total number of vertices is 5. Hence, any degree 4 vertex must have an edge with every other vertex. Since there are two degree 4 vertices, there cannot be a vertex with degree 1.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair DIS 2B

1 Short Answers

Note 5

In each part below, provide the number/equation and brief justification.

(a) A connected planar simple graph has 5 more edges than it has vertices. How many faces does it have?

(b) How many edges need to be removed from a 3-dimensional hypercube to get a tree?

(c) The Euler's formula $v - e + f = 2$ requires the planar graph to be connected. What is the analogous formula for planar graphs wth k connected components?

Solution:

(a) **7.**
Use Euler's formula $v + f = e + 2$.

(b) **5.**
The 3-dimensional hypercube has $3(2^3)/2 = 12$ edges and $2^3 = 8$ vertices. A tree on 8 vertices has 7 edges, so one needs to remove 5 edges.

(c) Let v_i, e_i, f_i be the number of vertices, edges, and faces respectively for the ith connected component. Let V, E, F be the analogous quantities for the entire graph. For each connected component i, Euler's equation gives $v_i - e_i + f_i = 2$. Summing over this for all i, we get

$$\sum_{i=1}^{k} (v_i - e_i + f_i) = 2k$$

When we add up the vertices and edges, the total count doesn't change. Only the number of faces changes when we consider multiple connected components. When we 'combine' two connected components (aka consider them to be the same graph), they end up sharing the 'infinite' face, so if we were to add the number of faces, we have to subtract 1 from our count. Thus,

$$V - E + \sum_{i=1}^{k} f_i = 2k$$

When we combine together k components, we end up overcounting the infinite face $k - 1$ times, so $F = \sum_i f_i - (k-1)$ Equivalently,

$$V - E + F + (k-1) = 2k$$

, or $V - E + F = k + 1$.

2 Always, Sometimes, or Never

Note 5

In each part below, you are given some information about a graph G. Using only the information in the current part, say whether G will always be planar, always be non-planar, or could be either. If you think it is always planar or always non-planar, prove it. If you think it could be either, give a planar example and a non-planar example.

(a) G can be vertex-colored with 4 colors.

(b) G requires 7 colors to be vertex-colored.

(c) $e \leq 3v - 6$, where e is the number of edges of G and v is the number of vertices of G.

(d) G is connected, and each vertex in G has degree at most 2.

(e) Each vertex in G has degree at most 2.

Solution:

(a) Either planar or non-planar. By the 4-color theorem, any planar graph can provide the planar example. The easiest non-planar example is $K_{3,3}$, which can be 2-colored because it is bipartite. (Certainly, any graph which can be colored using only 2 colors can also be colored using 4 colors.)

(b) Always non-planar. The 4-color theorem tells us that if a graph is planar, it can be colored using only 4 colors. The contrapositive of this is that if a graph requires more than 4 colors to vertex-color, it must be non-planar. (Using the 5- or 6-color theorem would also work.)

(c) Either planar or non-planar. From the notes, we know that every planar graph follows this formula, so any planar graph is a valid planar example. The easiest non-planar example is again $K_{3,3}$, which has $e = 9$ and $v = 6$, meaning our formula becomes $9 \leq 3(6) - 6 = 12$, which is certainly true.

(d) Always planar. There are two cases to deal with here: either G is a tree, or G is not a tree and so contains at least one cycle. In the former case, we're immediately done, since all trees are planar. In the latter case, consider any cycle in G. We know that every vertex in that cycle is adjacent to the vertex to its left in the cycle and to the vertex to its right in the cycle. But we also know that no vertex can be connected to more than two other vertices, so the cycle isn't connected to anything else. But G is a connected graph, so we must have that G is just a single large cycle. And we can certainly draw a simple cycle on a plane without crossing any edges, so even in this case G is still planar.

Alternatively, we can use Kuratowski's theorem; since each vertex has a degree of at most 2, it is impossible for G to contain K_5 or $K_{3,3}$. This means that G must be planar.

(e) Always planar. Each of G's connected components is connected and has no vertex of degree more than 2, so by the previous part, each of them must be planar. Thus, each of G's connected components must be planar, so G itself must be planar.

Alternatively, we can follow the same procedure as the previous alternate solution; each vertex still has a degree of at most 2, so it is impossible for G to contain K_5 or $K_{3,3}$. This means that G must be planar.

3 Graph Coloring

Note 5

Prove that a graph with maximum degree at most k is $(k+1)$-colorable.

Solution:

The natural way to try to prove this theorem is to use induction on the graph's maximum degree, k. Unfortunately, this approach is extremely difficult because covering all possible types of graphs when maximum degree changes requires extreme caution. You might be envisioning a certain graph as you write your proof, but your argument will likely not generalize. In graphs, typical good choices for the induction parameter are n, the number of nodes, or e, the number of edges. We typically shy away from inducting on degree.

We use induction on the number of vertices in the graph, which we denote by n. Let $P(n)$ be the proposition that an n-vertex graph with maximum degree at most k is $(k+1)$-colorable.

Base Case $n = 1$: A 1-vertex graph has maximum degree 0 and is 1-colorable, so $P(1)$ is true.

Inductive Step: Now assume that $P(n)$ is true, and let G be an $(n+1)$-vertex graph with maximum degree at most k. Remove a vertex v (and all edges incident to it), leaving an n-vertex subgraph,H. The maximum degree of H is at most k, and so H is $(k+1)$-colorable by our assumption $P(n)$. Now add back vertex v. We can assign v a color (from the set of $k+1$ colors) that is different from all its adjacent vertices, since there are at most k vertices adjacent to v and so at least one of the $k+1$ colors is still available. Therefore, G is $(k+1)$-colorable. This completes the inductive step, and the theorem follows by induction.

4 Hypercubes

Note 5

The vertex set of the n-dimensional hypercube $G = (V,E)$ is given by $V = \{0,1\}^n$ (recall that $\{0,1\}^n$ denotes the set of all n-bit strings). There is an edge between two vertices x and y if and only if x and y differ in exactly one bit position.

(a) Draw 1-, 2-, and 3-dimensional hypercubes and label the vertices using the corresponding bit strings.

(b) Show that the edges of an n-dimensional hypercube can be colored using n colors so that no pair of edges sharing a common vertex have the same color.

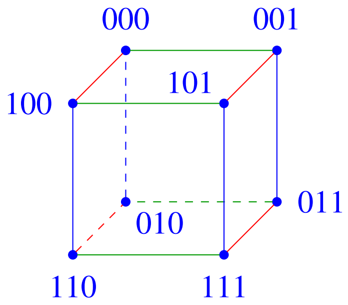
(c) Show that for any $n \geq 1$, the n-dimensional hypercube is bipartite.

Solution:

(a) The three hypercubes are a line, a square, and a cube, respectively. See also note 5 for pictures.

(b) Consider each edge that changes the ith bit for some $i \leq n$. Every vertex touches exactly one of these edges, because there is exactly one way to change the ith bit in any bitstring. Coloring each of these edges color i ensures that each vertex will then be adjacent to n differently colored edges, since there are n different bits to change, and no two edges representing bit changes on different bits have the same color.

An example for the three dimensional case is shown below (red is the first bit, blue is the second bit, and green is the third bit):



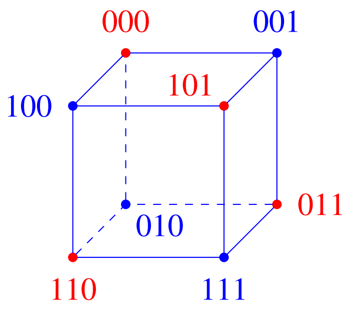
Alternate solution (using induction):

In the base case of $n = 1$, the hypercube of only one line can be edge colored with 1 color. Next, suppose that the n dimensional hypercube can be colored with n colors. Recall that the $n+1$ dimensional hypercube is composed of two n dimensional hypercubes; each of these hypercubes can be colored with n colors by the inductive hypothesis.

We can connect the two n dimensional hypercubes with edges colored with a different color; this will be our $(n+1)$th color. Since these new edges will always be between distinct pairs of vertices, one from each subcube, none of these new edges will share a vertex, giving a valid coloring of the $n+1$ dimensional hypercube with $n+1$ colors.

(c) Consider the vertices with an even number of 0 bits and the vertices with an odd number of 0 bits. Each vertex with an even number of 0 bits is adjacent only to vertices with an odd number of 0 bits, since each edge represents a single bit change (either a 0 bit is added by flipping a 1 bit, or a 0 bit is removed by flipping a 0 bit). Let L be the set of the vertices with an even number of 0 bits and let R be the vertices with an odd number of 0 bits, then no two adjacent vertices will belong to the same set.

An example for the three dimensional case is shown below (L are blue vertices, and R are red vertices):



Alternate solution (using induction and coloring):

It may be simpler to that a graph being 2-colorable is the same as being bipartite. Now, the argument is easier to state. First the base case is a hypercube with two vertices which is clearly two-colorable. Then notice, switching the colors in a two-coloring is still valid as if endpoints are differently colored, switching leaves them differently colored. Now, recursively one two colors the two subcubes the same, and then switches the colors in one subcube. The internal to subcube edges are fine by induction. The edges across are fine as the corresponding vertices are differently colored due to the switching.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair **DIS 3A**

1 Party Tricks

Note 6

You are at a party celebrating your completion of the CS 70 midterm. Show off your modular arithmetic skills and impress your friends by quickly figuring out the last digit(s) of each of the following numbers:

(a) Find the last digit of 11^{3142}.

(b) Find the last digit of 9^{9999}.

(c) Find the last digit of 3^{641}.

Solution:

(a) First, we notice that $11 \equiv 1 \pmod{10}$. So $11^{3142} \equiv 1^{3142} \equiv 1 \pmod{10}$, so the last digit is a 1.

(b) 9 is its own multiplicative inverse mod 10, so $9^2 \equiv 1 \pmod{10}$. Then

$$9^{9999} = 9^{2(4999)} \cdot 9 \equiv 1^{4999} \cdot 9 \equiv 9 \pmod{10},$$

so the last digit is a 9.

Another solution: We know $9 \equiv -1 \pmod{10}$, so

$$9^{9999} \equiv (-1)^{9999} \equiv -1 \equiv 9 \pmod{10}.$$

You could have also used this to say

$$9^{9999} \equiv (-1)^{9998} \cdot 9 \equiv 9 \pmod{10}.$$

(c) Notice that $3^4 = 9^2$ so using that $9^2 = 81 \equiv 1 \pmod{10}$, we have $3^4 \equiv 1 \pmod{10}$. We also have that $641 = 160 \cdot 4 + 1$, so

$$3^{641} \equiv 3^{4(160)} \cdot 3 \equiv 1^{160} \cdot 3 \equiv 3 \pmod{10},$$

making the last digit a 3.

2 Modular Potpourri

Note 6 Prove or disprove the following statements:

(a) There exists some $x \in \mathbb{Z}$ such that $x \equiv 3 \pmod{16}$ and $x \equiv 4 \pmod{6}$.

(b) $2x \equiv 4 \pmod{12} \iff x \equiv 2 \pmod{12}$.

(c) $2x \equiv 4 \pmod{12} \iff x \equiv 2 \pmod{6}$.

Solution:

(a) Impossible.

Suppose there exists an x satisfying both equations.

From $x \equiv 3 \pmod{16}$, we have $x = 3 + 16k$ for some integer k. This implies $x \equiv 1 \pmod{2}$.

From $x \equiv 4 \pmod{6}$, we have $x = 4 + 6l$ for some integer l. This implies $x \equiv 0 \pmod{2}$.

Now we have $x \equiv 1 \pmod{2}$ and $x \equiv 0 \pmod{2}$. Contradiction.

(b) False, consider $x \equiv 8 \pmod{12}$.

The reason we can't eliminate the 2 in the first equation to get the second equation is because 2 does not have a multiplicative inverse modulo 12, as 2 and 12 are not coprime.

(c) True. We can write $2x \equiv 4 \pmod{12}$ as $2x = 4 + 12k$ for some $k \in \mathbb{Z}$. Dividing by 2, we have $x = 2 + 6k$ for the same $k \in \mathbb{Z}$. This is equivalent to saying $x \equiv 2 \pmod{6}$.

3 Modular Inverses

Note 6 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) For all $n \in \mathbb{N}$, is $3 + 14n$ an inverse of 5 modulo 14?

(d) Does 4 have an 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

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

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

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

4 Fibonacci GCD

Note 6

The Fibonacci sequence is given by $F_n = F_{n-1} + F_{n-2}$, where $F_0 = 0$ and $F_1 = 1$. Prove that, for all $n \geq 1$, $\gcd(F_n, F_{n-1}) = 1$.

Solution:

Proceed by induction.

Base Case: We have $\gcd(F_1, F_0) = \gcd(1, 0) = 1$, which is true.

Inductive Hypothesis: Assume we have $\gcd(F_k, F_{k-1}) = 1$ for some $k \geq 1$.

Inductive Step: Now we need to show that $\gcd(F_{k+1}, F_k) = 1$ as well.

We can show that:

$$\gcd(F_{k+1}, F_k) = \gcd(F_k + F_{k-1}, F_k) = \gcd(F_k, F_{k-1}) = 1.$$

Note that the second expression comes from the definition of Fibonacci numbers. The last expression comes from Euclid's GCD algorithm, in which $\gcd(x, y) = \gcd(y, x \bmod y)$, since

$$F_k + F_{k-1} \equiv F_{k-1} \pmod{F_k}.$$

Therefore the statement is also true for $n = k+1$.

By the rule of induction, we can conclude that $\gcd(F_n, F_{n-1}) = 1$ for all $n \geq 1$, where $F_0 = 0$ and $F_1 = 1$ and $F_n = F_{n-1} + F_{n-2}$.

CS 70 Spring 2024 — Discrete Mathematics and Probability Theory — Seshia, Sinclair — DIS 3B

1 Extended Euclid: Two Ways

Note 6

In this problem, we will explore the Extended Euclid's Algorithm: first, the traditional implementation, and second, a faster, iterative version. Both ways yield the same result.

Parts (b) and (c) explore the traditional Extended Euclid's Algorithm. The bolded numbers below keep track of which numbers appeared as inputs to the gcd call. Remember that we are interested in writing the GCD as a linear combination of the original inputs, so we don't want to accidentally simplify the expressions and eliminate the inputs.

(a) As motivation, suppose we've found values of a and b such that $54a + 17b = 1$. With this knowledge, what is $17^{-1} \pmod{54}$?

(b) Note that $x \bmod y$, by definition, is always x minus a multiple of y. So, in the execution of Euclid's algorithm, each newly introduced value can always be expressed as a "combination" of the previous two, like so:

$$\begin{aligned}
\gcd(54,17) &= \gcd(17,3) & \mathbf{3} &= 1\times\mathbf{54} - 3\times\mathbf{17}\\
&= \gcd(3,2) & \mathbf{2} &= 1\times\mathbf{17} - \underline{\hspace{2em}}\times\mathbf{3}\\
&= \gcd(2,1) & \mathbf{1} &= 1\times\mathbf{3} - \underline{\hspace{2em}}\times\mathbf{2}\\
&= \gcd(1,0) & [\mathbf{0} &= 1\times\mathbf{2} - \underline{\hspace{2em}}\times\mathbf{1}]\\
&= 1.
\end{aligned}$$

(Fill in the blanks)

(c) Recall that our goal is to fill out the blanks in

$$1 = \underline{\hspace{2em}}\times\mathbf{54} + \underline{\hspace{2em}}\times\mathbf{17}.$$

To do so, we work back up from the bottom, and express the gcd above as a combination of the two arguments on each of the previous lines:

$$\begin{aligned}
1 &= \underline{\hspace{2em}}\times\mathbf{3} + \underline{\hspace{2em}}\times\mathbf{2}\\
&=\\
&= \underline{\hspace{2em}}\times\mathbf{17} + \underline{\hspace{2em}}\times\mathbf{3}\\
&=\\
&= \underline{\hspace{2em}}\times\mathbf{54} + \underline{\hspace{2em}}\times\mathbf{17}
\end{aligned}$$

What does this imply, in this case, about the multiplicative inverse of 17, in arithmetic mod 54?

(d) In the previous parts, we used a recursive method to write $\gcd(54,17)$ as a linear combination of 54 and 17. We can also compute the same result iteratively—this is an alternative to the above method that is oftentimes faster. We begin by writing equations for our initial arguments, 54 and 17, as a linear combination of themselves:

$$54 = 1 \times \mathbf{54} + 0 \times \mathbf{17} \qquad (E_1)$$
$$17 = 0 \times \mathbf{54} + 1 \times \mathbf{17} \qquad (E_2)$$

We can then use these initial equations (labeled E_1 and E_2 for ease of reference) to iteratively write reduced values as linear combinations of 54 and 17, until we are able to write an equation for $\gcd(54,17)$, as desired.

In particular, we want to subtract as many multiples of the second equation as possible from the first to create a new equation with a lower LHS value. We can keep iterating until the LHS becomes $\gcd(54,17) = 1$.

$$\underline{\qquad} = \underline{\qquad} \times \mathbf{54} + \underline{\qquad} \times \mathbf{17} \qquad (E_3 = E_1 - \underline{\qquad} \times E_2)$$
$$\underline{\qquad} = \underline{\qquad} \times \mathbf{54} + \underline{\qquad} \times \mathbf{17} \qquad (E_4 = E_2 - \underline{\qquad} \times E_3)$$
$$1 = \underline{\qquad} \times \mathbf{54} + \underline{\qquad} \times \mathbf{17} \qquad (E_5 = E_3 - \underline{\qquad} \times E_4)$$

What does this imply, in this case, about the multiplicative inverse of 17, in arithmetic mod 54? Verify that your answer is equivalent to the previous part.

(e) Calculate the gcd of 17 and 39, and determine how to express this as a "combination" of 17 and 39. What does this imply, in this case, about the multiplicative inverse of 17, in arithmetic mod 39?

Solution:

(a) If we take the equation $54a + 17b = 1 \pmod{54}$, the first term goes to zero (as it is a multiple of 54). This means that we're left with $17b \equiv 1 \pmod{54}$, giving us that $17^{-1} \equiv b \pmod{54}$.

In other words, the coefficients we get from the extended Euclidean algorithm give us the multiplicative inverse directly. This is one of the main reasons why the extended Euclidean algorithm is useful.

(b) Filling in the blanks,

$$\mathbf{3} = 1 \times \mathbf{54} - 3 \times \mathbf{17}$$
$$\mathbf{2} = 1 \times \mathbf{17} - 5 \times \mathbf{3}$$
$$\mathbf{1} = 1 \times \mathbf{3} - 1 \times \mathbf{2}$$
$$[\mathbf{0} = 1 \times \mathbf{2} - 2 \times \mathbf{1}]$$

It may be easier to think about this in a rearranged form: $\mathbf{54} = 3 \times \mathbf{17} + \mathbf{3}$, etc.; this directly corresponds to the 54 mod $17 = 3$ operation in the forward pass, and the desired blank comes from $\lfloor 54/17 \rfloor$.

(c) Working our way backward up the equalities and substituting them in, we have

$$\begin{aligned}
1 &= 1 \times \mathbf{3} - 1 \times \mathbf{2} \\
&= 1 \times \mathbf{3} - 1 \times (1 \times \mathbf{17} - 5 \times \mathbf{3}) \\
&= -1 \times \mathbf{17} + 6 \times \mathbf{3} \\
&= -1 \times \mathbf{17} + 6 \times (1 \times \mathbf{54} - 3 \times \mathbf{17}) \\
&= 6 \times \mathbf{54} - 19 \times \mathbf{17}
\end{aligned}$$

We get that the multiplicative inverse of 17 mod 54 is -19, or 35. Note that $-19 \equiv 35 \bmod 54$.

(d) We have the following operations on the equations:

$$\begin{aligned}
54 &= 1 \times \mathbf{54} + 0 \times \mathbf{17} && (E_1) \\
17 &= 0 \times \mathbf{54} + 1 \times \mathbf{17} && (E_2) \\
3 &= 1 \times \mathbf{54} - 3 \times \mathbf{17} && (E_3 = E_1 - 3E_2) \\
2 &= -5 \times \mathbf{54} + 16 \times \mathbf{17} && (E_4 = E_2 - 5E_3) \\
1 &= 6 \times \mathbf{54} - 19 \times \mathbf{17} && (E_5 = E_3 - E_4)
\end{aligned}$$

Notice that the LHS also corresponds to the simplifications in the forward pass of the Euclidean algorithm; we're doing the same calculations (i.e. to determine how much to subtract), and we're also doing the backward pass at the same time. This is why the iterative method can be more intuitive and quicker than the recursive method in the previous parts.

Again, we get that the multiplicative inverse of 17 mod 54 is -19, or 35.

(e) With the recursive algorithm, we have

$$\begin{aligned}
\gcd(39,17) &= \gcd(17,5) && \mathbf{5} = 1 \times \mathbf{39} - 2 \times \mathbf{17} \\
&= \gcd(5,2) && \mathbf{2} = 1 \times \mathbf{17} - 3 \times \mathbf{5} \\
&= \gcd(2,1) && \mathbf{1} = 1 \times \mathbf{5} - 2 \times \mathbf{2} \\
&= \gcd(1,0) && [\mathbf{0} = 1 \times \mathbf{2} - 2 \times \mathbf{1}]
\end{aligned}$$

Going back up, we have

$$\begin{aligned}
\mathbf{1} &= 1 \times \mathbf{5} - 2 \times \mathbf{2} \\
&= 1 \times \mathbf{5} - 2 \times (1 \times \mathbf{17} - 3 \times \mathbf{5}) \\
&= -2 \times \mathbf{17} + 7 \times \mathbf{5} \\
&= -2 \times \mathbf{17} + 7 \times (1 \times \mathbf{39} - 2 \times \mathbf{17}) \\
&= 7 \times \mathbf{39} - 16 \times \mathbf{17}
\end{aligned}$$

This leaves us with a final answer of $1 = 7 \times \mathbf{39} - 16 \times \mathbf{17}$, making the inverse $17^{-1} \equiv -16 \equiv 23 \pmod{39}$.

With the iterative algorithm, we have

$$\begin{aligned}
39 &= 1 \times \mathbf{39} + 0 \times \mathbf{17} && (E_1)\\
17 &= 0 \times \mathbf{39} + 1 \times \mathbf{17} && (E_2)\\
5 &= 1 \times \mathbf{39} - 2 \times \mathbf{17} && (E_3 = E_1 - 2E_2)\\
2 &= -3 \times \mathbf{39} + 7 \times \mathbf{17} && (E_4 = E_2 - 3E_3)\\
1 &= 7 \times \mathbf{39} - 16 \times \mathbf{17} && (E_5 = E_3 - 2E_4)
\end{aligned}$$

2 Chinese Remainder Theorem Practice

Note 6

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

$$\begin{aligned}
x &\equiv 1 \pmod 3\\
x &\equiv 3 \pmod 7\\
x &\equiv 4 \pmod{11}
\end{aligned} \tag{1}$$

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

$$a \equiv 1 \pmod 3\ ;\ a \equiv 0 \pmod 7\ ;\ a \equiv 0 \pmod{11}, \tag{2}$$
$$b \equiv 0 \pmod 3\ ;\ b \equiv 1 \pmod 7\ ;\ b \equiv 0 \pmod{11}, \tag{3}$$
$$c \equiv 0 \pmod 3\ ;\ c \equiv 0 \pmod 7\ ;\ c \equiv 1 \pmod{11}. \tag{4}$$

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 satisfies (1).

(b) Find a natural number a that satisfies (2). That is, $a \equiv 1 \pmod 3$ and is a multiple of 7 and 11.

It may help to start with a number that is a multiple of both 7 and 11; what number should we multiply this by in order to make it equivalent to $1 \pmod 3$?

(c) Find a natural number b that satisfies (3). That is, $b \equiv 1 \pmod 7$ and is a multiple of 3 and 11.

(d) Find a natural number c that satisfies (4). That is, $c \equiv 1 \pmod{11}$ and is a multiple of 3 and 7.

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

Solution:

(a) Observe that $a+3b+4c \equiv 1+0+0 \pmod 3$, $a+3b+4c \equiv 0+3+0 \pmod 7$ and $a+3b+4c \equiv 0+0+4 \pmod{11}$. Therefore $x = a+3b+4c$ indeed satisfies the conditions in (1).

(b) This question asks to find a number $0 \leq a < 3 \times 7 \times 11$ that is divisible by 7 and 11 and has a remainder of 1 when divided by 3.

Starting with a number divisible by 7 and 11, we can start with $7 \cdot 11 = 77$. Notice that we can multiply by the multiplicative inverse mod 3 to make it equivalent to $1 \pmod 3$. In particular, since $77 \cdot 77^{-1} \equiv 1 \pmod 3$, we just need to compute

$$77^{-1} \equiv 2^{-1} \equiv 2 \pmod 3.$$

This gives us $a = 77 \cdot 2 = 154$.

We can check to make sure that what we've computed actually satisfies (2):

$$\begin{aligned}
154 &= 3 \cdot 51 + 1 \equiv 1 \pmod 3 \\
154 &= 22 \cdot 7 \equiv 0 \pmod 7 \\
154 &= 14 \cdot 11 \equiv 0 \pmod{11}
\end{aligned}$$

Taking a step back, notice that what we've computed is

$$a = (7 \cdot 11) \cdot \big((7 \cdot 11)^{-1} \bmod 3\big).$$

Here, the first term ensures that we have a multiple of 7 and 11, and the last term ensures that we have a quantity equivalent to $1 \pmod 3$.

(c) Using a similar approach here, we can start with a multiple of 3 and 11; namely, $3 \cdot 11 = 33$.

Here, we can multiply by its multiplicative inverse mod 7 to make it equivalent to $1 \pmod 7$. In particular, we just need to compute

$$33^{-1} \equiv 5^{-1} \equiv 3 \pmod 7.$$

This gives us $b = 33 \cdot 3 = 99$.

Again, notice that we've essentially just computed

$$b = (3 \cdot 11) \cdot \big((3 \cdot 11)^{-1} \bmod 7\big).$$

(d) Similarly, we can start with a multiple of 3 and 7; namely, $3 \cdot 7 = 21$.

Here, we can multiply by its multiplicative inverse mod 11 to make it equivalent to $1 \pmod{11}$. In particular, we just need to compute

$$21^{-1} \equiv 10^{-1} \equiv 10 \pmod{11}.$$

This gives us $c = 21 \cdot 10 = 210$.

Again, notice that we've essentially just computed

$$c = (3 \cdot 7) \cdot \big((3 \cdot 7)^{-1} \bmod 11\big).$$

(e) From Parts (b), (c) and (d) we've found $a = 154$, $b = 99$, and $c = 210$ which satisfies (2), (3) and (4) respectively. Together with Part (a) of the question, this implies that

$$x = a + 3b + 4c = 154 + 3 \cdot 99 + 4 \cdot 210 = 154 + 297 + 840 = 1291$$

satisfies the required criterion in (1).

To verify this, observe that

$$\begin{aligned} 1291 &= 430 \times 3 + 1 \equiv 1 \pmod{3} \\ 1291 &= 184 \times 7 + 3 \equiv 3 \pmod{7} \\ 1291 &= 117 \times 11 + 4 \equiv 4 \pmod{11} \end{aligned}$$

Further, this solution will be unique mod $3 \cdot 7 \cdot 11 = 231$, so we have $x \equiv 1291 \equiv 136 \pmod{231}$.

As a side note, what we're essentially doing here is computing values that satisfy exactly one of the equivalences, while not affecting any of the other equivalences. In particular, suppose we have a system of k modular equations $x \equiv a_i \pmod{m_i}$ for $i = 1$ through k. For each equation, we want a value $b_i \equiv 1 \pmod{m_i}$ and $b_i \equiv 0 \pmod{m_j}$ for $j \neq i$, such that $a_i b_i$ satisfies exactly the mod m_i equivalence but is equivalent to zero for everything else. This way, adding up all of the $a_i b_i$'s will give us a quantity that satisfies all of the equivalences.

Computing each b_i can be written as the following formula:

$$b_i = \frac{M}{m_i} \cdot \left(\left(\frac{M}{m_i} \right)^{-1} \bmod m_i \right),$$

where $M = m_1 \cdot m_2 \cdots m_k$. The first term ensures that $b_i \equiv 0 \pmod{m_j}$ for $j \neq i$, and the second term ensures that $b_i \equiv 1 \pmod{m_i}$. The solution can then be computed by

$$x \equiv \sum_{i=1}^{k} a_i b_i \pmod{M}.$$

3 Baby Fermat

Note 6

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 infinite 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 is the value of $a^{i-j} \pmod{m}$?

(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
Spring 2024 Seshia, Sinclair
DIS 4A

1 RSA Warm-Up

Note 7

Consider an RSA scheme with modulus $N = pq$, where p and q are distinct prime numbers larger than 3.

(a) What is wrong with using the exponent $e = 2$ in an RSA public key?

(b) Recall that e must be relatively prime to $p-1$ and $q-1$. Find a condition on p and q such that $e = 3$ is a valid exponent.

(c) Now suppose that $p = 5$, $q = 17$, and $e = 3$. What is the public key?

(d) What is the private key?

(e) Alice wants to send a message $x = 10$ to Bob. What is the encrypted message $E(x)$ she sends using the public key?

(f) Suppose Bob receives the message $y = 19$ from Alice. What equation would he use to decrypt the message? What is the decrypted message?

Solution:

(a) To find the private key d from the public key (N, e), we need $\gcd(e,(p-1)(q-1)) = 1$. However, $(p-1)(q-1)$ is necessarily even since p, q are distinct odd primes, so if $e = 2$, $\gcd(e,(p-1)(q-1)) = 2$, and a private key does not exist. (Note that this shows that e should more generally never be even.)

(b) Both p and q must be of the form $3k+2$. $p = 3k+1$ is a problem since then $p-1$ has a factor of 3 in it. $p = 3k$ is a problem because then p is not prime.

(c) $N = p \cdot q = 85$ and $e = 3$ are displayed publicly. Note that in practice, p and q should be much larger (512-bit) numbers. We are only choosing small numbers here to allow manual computation.

(d) We must have $ed = 3d \equiv 1 \pmod{64}$, so $d = 43$. Reminder: we would do this by using extended gcd with $x = 64$ and $y = 3$. We get $\gcd(x,y) = 1 = ax + by$, and $a = 1$, $b = -21$.

(e) We have $E(x) = x^3 \pmod{85}$, where $E(x)$ is the encryption function. $10^3 \equiv 65 \pmod{85}$, so $E(x) = 65$.

(f) We have $D(y) = y^{43} \pmod{85}$, where $D(y)$ is the decryption function, the inverse of $E(x)$.

$$x \equiv 19^{43} \pmod{85}$$

From CRT we know that for coprime numbers p and q if

$$\begin{aligned} x &\equiv a \pmod{p} \\ x &\equiv b \pmod{q} \end{aligned}$$

then

$$x = aqq_1 + bpp_1 \pmod{pq}$$

where $p_1 = p^{-1} \pmod{q}$ and $q_1 = q^{-1} \pmod{p}$.
In our case we have $p = 5$ and $q = 17$. So

$$x \equiv 19^{43} \equiv (-1)^{43} \equiv -1 \equiv 4 \pmod{5}$$

and

$$\begin{aligned} x &\equiv 19^{43} \pmod{17} \\ x &\equiv (2)^{43} \pmod{17} \\ x &\equiv (2^4)^{10} \cdot 2^3 \pmod{17} \\ x &\equiv (-1)^{10} \cdot 8 \pmod{17} \\ x &\equiv 8 \pmod{17} \end{aligned}$$

Hence

$$x = a = 4 \pmod{5} \qquad x = b = 8 \pmod{17}$$

and

$$\begin{aligned} p_1 &= p^{-1} \pmod{17} = 5^{-1} \pmod{17} = 7 \\ q_1 &= q^{-1} \pmod{5} = 17^{-1} \pmod{5} = 3 \end{aligned}$$

So we have

$$\begin{aligned} x &\equiv aqq_1 + bpp_1 \pmod{pq} \\ x &\equiv 4 \cdot 17 \cdot 3 + 8 \cdot 5 \cdot 7 \pmod{85} \\ x &\equiv 4 \cdot 17 \cdot 3 + 280 \pmod{85} \\ x &\equiv 17 \cdot (12) + 280 \pmod{85} \\ x &\equiv 17 \cdot (10 + 2) + 280 \pmod{85} \\ x &\equiv 34 + 25 \pmod{85} \\ x &\equiv 59 \pmod{85} \end{aligned}$$

so $D(y) = 59$.

2 RSA with Multiple Keys

Note 7

Members of a secret society know a secret word. They transmit this secret word x between each other many times, each time encrypting it with the RSA method. Eve, who is listening to all of their communications, notices that in all of the public keys they use, the exponent e is the same. Therefore the public keys used look like $(N_1, e), \dots, (N_k, e)$ where no two N_i's are the same. Assume that the message is x such that $0 \le x < N_i$ for every i.

Further, in all of the subparts, you may assume that Eve knows the details of the modified RSA schemes (i.e. Eve knows the format of the N_i's, but not the specific values used to compute the N_i's).

(a) Suppose Eve sees the public keys $(p_1q_1, 7)$ and $(p_1q_2, 7)$ as well as the corresponding transmissions. Can Eve use this knowledge to break the encryption? If so, how? Assume that Eve cannot compute prime factors efficiently. Think of p_1, q_1, q_2 as massive 1024-bit numbers. Assume p_1, q_1, q_2 are all distinct and are valid primes for RSA to be carried out.

(b) The secret society has wised up to Eve and changed their choices of N, in addition to changing their word x. Now, Eve sees keys $(p_1q_1, 3)$, $(p_2q_2, 3)$, and $(p_3q_3, 3)$ along with their transmissions. Argue why Eve cannot break the encryption in the same way as above. Assume $p_1, p_2, p_3, q_1, q_2, q_3$ are all distinct and are valid primes for RSA to be carried out.

(c) Let's say the secret x was not changed ($e = 3$), so they used the same public keys as before, but did not transmit different messages. How can Eve figure out x?

Solution:

(a) Normally, the difficulty of cracking RSA hinges upon the believed difficulty of factoring large numbers. If Eve were given just p_1q_1, she would (probably) not be able to figure out the factors.

However, Eve has access to two public keys, so yes, she will be able to figure it out. Note that $\gcd(p_1q_1, p_1q_2) = p_1$. Taking GCDs is actually an efficient operation thanks to the Euclidean Algorithm. Therefore, she can figure out the value of p_1, and from there figure out the value of q_1 and q_2 since she has p_1q_1 and p_1q_2.

(b) Since none of the N's have common factors, she cannot find a GCD to divide out of any of the Ns. Hence the approach above does not work.

(c) Eve observes $x^3 \pmod{N_1}$, $x^3 \pmod{N_2}$, $x^3 \pmod{N_3}$. Since all N_1, N_2, N_3 are pairwise relatively prime, Eve can use the Chinese Remainder Theorem to figure out $x^3 \pmod{N_1N_2N_3}$. However, once she gets that, she knows x, since $x < N_1$, $x < N_2$, and $x < N_3$, which implies $x^3 < N_1N_2N_3$, so she can directly take the cube root of the result from CRT. Uh oh!

3 RSA for Concert Tickets

Note 7

Alice wants to tell Bob her concert ticket number, m, which is an integer between 0 and 100 inclusive. She wants to tell Bob over an insecure channel that Eve can listen in on, but Alice does not want Eve to know her ticket number.

(a) Bob announces his public key $(N = pq, e)$, where N is large (512 bits). Alice encrypts her message using RSA. Eve sees the encrypted message, and figures out what Alice's ticket number is. How did she do it?

(b) Alice decides to be a bit more elaborate. She picks a random number r that is 256 bits long, so that it is too hard to guess. She encrypts that and sends it to Bob, and also computes rm, encrypts that, and sends it to Bob. Eve is aware of what Alice did, but does not know the value of r. How can she figure out m? (You may assume that r is coprime to N.)

Solution:

(a) There are only 101 possible values for Alice's ticket number, so Eve can try encrypting all 101 values with Bob's public key and find out which one matches the one Alice sent.

(b) Alice sends $x = r^e \pmod{pq}$, as well as $y = (rm)^e = r^e m^e = xm^e \pmod{pq}$. We can find $x^{-1} \pmod{N}$ using the Extended Euclidean Algorithm, and multiplying this value by y gives us $m^e \pmod{N}$. Now we proceed as in the previous part to find m.

Another approach is to compute xm^e for all 101 values of m, and compare the value to y, checking which one matches.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 4B

1 Polynomial Practice

Note 8

(a) If f and g are non-zero real polynomials, how many real 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 $\mathrm{GF}(p)$.

(i) We say a polynomial $f = 0$ if $\forall x, f(x) = 0$. Show that if $f \cdot g = 0$, it is not always 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 $\mathrm{GF}(5)$ that satisfies $f(0) = 1, f(2) = 2, f(4) = 0$. How many such polynomials of degree at most 4 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. 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) There are a couple counterexamples:

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 FLT, $x^{p-1}-1$ 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 $S = \{0, \dots, p-1\}$. 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) 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 Lagrange Interpolation in Finite Fields

Note 8

Find a unique polynomial $p(x)$ of degree at most 2 that passes through points $(-1,3)$, $(0,1)$, and $(1,2)$ in modulo 5 arithmetic using the Lagrange interpolation.

(a) Find $p_{-1}(x)$ where $p_{-1}(0) \equiv p_{-1}(1) \equiv 0 \pmod 5$ and $p_{-1}(-1) \equiv 1 \pmod 5$.

(b) Find $p_0(x)$ where $p_0(-1) \equiv p_0(1) \equiv 0 \pmod 5$ and $p_0(0) \equiv 1 \pmod 5$.

(c) Find $p_1(x)$ where $p_1(-1) \equiv p_1(0) \equiv 0 \pmod 5$ and $p_1(1) \equiv 1 \pmod 5$.

(d) Construct $p(x)$ using a linear combination of $p_{-1}(x)$, $p_0(x)$, and $p_1(x)$.

Solution:

(a) We see

$$\begin{aligned} p_{-1}(x) &\equiv (x-0)(x-1)\big((-1-0)(-1-1)\big)^{-1} \\ &\equiv (2)^{-1} x(x-1) \pmod 5 \\ &\equiv 3x(x-1) \pmod 5. \end{aligned}$$

(b) We see

$$\begin{aligned} p_0(x) &\equiv (x+1)(x-1)\big((0+1)(0-1)\big)^{-1} \\ &\equiv (-1)^{-1}(x-1)(x+1) \pmod 5 \\ &\equiv 4(x-1)(x+1) \pmod 5. \end{aligned}$$

(c) We see

$$\begin{aligned} p_1(x) &\equiv (x+1)(x-0)\big((1+1)(1-0)\big)^{-1} \\ &\equiv (2)^{-1}x(x+1) \pmod 5 \\ &\equiv 3x(x+1) \pmod 5. \end{aligned}$$

(d) Putting everything together,

$$\begin{aligned} p(x) &= 3p_{-1}(x) + 1p_0(x) + 2p_1(x) \\ &= 9x(x-1) + 4(x-1)(x+1) + 6x(x+1) \\ &\equiv 4x^2 - 3x - 4 \pmod 5 \\ &\equiv 4x^2 + 2x + 1 \pmod 5. \end{aligned}$$

3 Secrets in the United Nations

Note 8

A vault in the United Nations can be opened with a secret combination $s \in \mathbb{Z}$. In only two situations should this vault be opened: (i) all 193 member countries must agree, or (ii) at least 55 countries, plus the U.N. Secretary-General, must agree.

(a) Propose a scheme that gives private information to the Secretary-General and all 193 member countries so that the secret combination s can only be recovered under either one of the two specified conditions.

(b) The General Assembly of the UN decides to add an extra level of security: each of the 193 member countries has a delegation of 12 representatives, all of whom must agree in order for that country to help open the vault. Propose a scheme that adds this new feature. The scheme should give private information to the Secretary-General and to each representative of each country.

Solution:

(a) Create a polynomial of degree 192 and give each country one point. Give the Secretary General $193-55 = 138$ distinct points, so that if she collaborates with 55 countries, they will have a total of 193 points and can reconstruct the polynomial. Without the Secretary-General, the polynomial can still be recovered if all 193 countries come together. (We do all our work in $\mathrm{GF}(p)$ where $p \geq d+1$).

Alternatively, we could have one scheme for condition (i) and another for (ii). The first condition is the secret-sharing setup we discussed in the notes, so a single polynomial of degree 192 suffices, with each country receiving one point, and evaluation at zero returning the combination s. For the second condition, create a polynomial f of degree 1 with $f(0) = s$, and give $f(1)$ to the Secretary-General. Now create a second polynomial g of degree 54, with $g(0) = f(2)$, and give one point of g to each country. This way any 55 countries can recover $g(0) = f(2)$, and then can consult with the Secretary-General to recover $s = f(0)$ from $f(1)$ and $f(2)$.

(b) We'll layer an *additional* round of secret-sharing onto the scheme from part (a). If t_i is the key given to the ith country, produce a degree-11 polynomial f_i so that $f_i(0) = t_i$, and give one point of f_i to each of the 12 delegates. Do the same for each country (using different f_i each time, of course).

4 To The Moon!

Note 8

A secret number s is required to launch a rocket, and Alice distributed the values $(1, p(1)), (2, p(2)), \ldots, (n+1, p(n+1))$ of a degree n polynomial p to a group of \$GME holders $\text{Bob}_1, \ldots, \text{Bob}_{n+1}$. As usual, she chose p such that $p(0) = s$. Bob_1 through Bob_{n+1} now gather to jointly discover the secret. However, Bob_1 is secretly a partner at Melvin Capital and already knows s, and wants to sabotage $\text{Bob}_2, \ldots, \text{Bob}_{n+1}$, 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 terms variables known in the problem, such as s', s or y_1) 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
Spring 2024 Seshia, Sinclair
DIS 5A

1 Berlekamp-Welch Warm Up

Note 8
Note 9

Let $P(i)$, a polynomial applied to the input i, be the original encoded polynomial before sent, and let r_i be the received info for the input i which may or may not be corrupted.

(a) If you want to send a length-n message, what should the degree of $P(x)$ be? Why?

(b) When does $r_i = P(i)$? When does r_i not equal $P(i)$?

(c) If there are at most k erasure errors, how many packets should you send? If there are at most k general errors, how many packets should you send? (We will see the reason for this later.) Now we will only consider general errors.

(d) What do the roots of the error polynomial $E(x)$ represent? Does the receiver know the roots of $E(x)$? If there are at most k errors, what is the maximum degree of $E(x)$? Using the information about the degree of $P(x)$ and $E(x)$, what is the degree of $Q(x) = P(x)E(x)$?

(e) Why is the equation $Q(i) = P(i)E(i) = r_iE(i)$ always true? (Consider what happens when $P(i) = r_i$, and what happens when $P(i)$ does not equal r_i.)

(f) In the polynomials $Q(x)$ and $E(x)$, how many total unknown coefficients are there? (These are the variables you must solve for. Think about the degree of the polynomials.) When you receive packets, how many equations do you have? Do you have enough equations to solve for all of the unknowns? (Think about the answer to the earlier question - does it make sense now why we send as many packets as we do?)

(g) If you have $Q(x)$ and $E(x)$, how does one recover $P(x)$? If you know $P(x)$, how can you recover the original message?

Solution:

(a) P has degree $n-1$ since n points would determine a degree $n-1$ polynomial.

(b) $r_i = P(i)$ when the received packet is correct. r_i does not equal $P(i)$ the received packet is corrupted.

(c) We send $n+k$ packets when we have k erasures and $n+2k$ packets for k general errors.

(d) The roots of error polynomial $E(x)$ represent the locations of corrupted packets. The receiver does not know the roots of $E(x)$. $E(x)$ is a polynomial that the receiver needs to compute in order to obtain $P(x)$. If there are at most k errors, then the maximum degree of $E(x)$ is k. The maximum degree of Q is $(n-1)+(k)=n+k-1$ since the degree of P is $n-1$ and the degree of E is at most k.

(e) If there is no error at point i, $P(i)=r_i$ and then multiplying each side by $E(i)$ gives $P(i)E(i)=r_iE(i)$. If there is an error at point i, then $E(i)=0$, which means $P(i)E(i)=r_iE(i)=0$.

(f) The maximum degree of $Q(x)$ is $n+k-1$, so the number of unknowns is $n+k$. The maximum degree of $E(x)$ is k, which would mean there would be $k+1$ unknowns. However, we know that the coefficient of x^k is 1 in $E(x)$, so the number of unknowns is k.

The total number of unknowns is $(n+k)+(k)=n+2k$

There are $n+2k$ equations, which is enough to solve for $n+2k$ unknowns.

(g) We can compute $P(x)$ using the equation: $P(x)=Q(x)/E(x)$. To recover the message, we compute $P(i)$ for $1\le i\le n$.

2 Berlekamp-Welch Algorithm

Note 8
Note 9

In this question we will send the message $(m_0,m_1,m_2)=(1,1,4)$ of length $n=3$. We will use an error-correcting code for $k=1$ general error, doing arithmetic over GF(5).

(a) Construct a polynomial $P(x) \pmod 5$ of degree at most 2, so that

$$P(0)=1, \qquad P(1)=1, \qquad P(2)=4.$$

What is the message (c_0,c_1,c_2,c_3,c_4) that is sent?

(b) Suppose the message is corrupted by changing c_0 to 0. Set up the system of linear equations in the Berlekamp-Welch algorithm to find $Q(x)$ and $E(x)$.

(c) Assume that after solving the equations in part (b) we get $Q(x)=4x^3+x^2+x$ and $E(x)=x$. Show how to recover the original message from Q and E.

Solution:

(a) We use Lagrange interpolation to construct the unique quadratic polynomial $P(x)$ such that $P(0)=m_0=1, P(1)=m_1=1, P(2)=m_2=4$. Doing all arithmetic over GF(5), so that i.e.

$2^{-1} = 3 \pmod 5$,

$$\begin{aligned}
\Delta_0(x) &= \frac{(x-1)(x-2)}{(0-1)(0-2)} = \frac{x^2-3x+2}{2} \equiv 3(x^2-3x+2) \pmod 5 \\
\Delta_1(x) &= \frac{(x-0)(x-2)}{(1-0)(1-2)} = \frac{x^2-2x}{-1} \equiv 4(x^2-2x) \pmod 5 \\
\Delta_2(x) &= \frac{(x-0)(x-1)}{(2-0)(2-1)} = \frac{x^2-x}{2} \equiv 3(x^2-x) \pmod 5 \\
P(x) &= m_0\Delta_0(x) + m_1\Delta_1(x) + m_2\Delta_2(x) \\
&= 1\Delta_0(x) + 1\Delta_1(x) + 4\Delta_2(x) \\
&\equiv 4x^2 + x + 1 \pmod 5
\end{aligned}$$

For the final message we need to add 2 redundant points of P. Since 3 and 4 are the only points in GF(5) that we have not used yet, we compute $P(3) = 0, P(4) = 4$, and so our message is $(1,1,4,0,4)$.

(b) The message received is $(c'_0, c'_1, c'_2, c'_3, c'_4) = (0,1,4,0,4)$. Let $R(x)$ be the function such $R(i) = c'_i$ for $0 \le i < 5$. Let $E(x) = x + b_0$ be the error-locator polynomial, and $Q(x) = P(x)E(x) = a_3x^3 + a_2x^2 + a_1x + a_0$. Since $Q(i) = P(i)E(i) = R(i)E(i)$ for $0 \le i < 5$, we have the following equalities $\pmod 5$

$$\begin{aligned}
Q(0) &= 0E(0) \\
Q(1) &= 1E(1) \\
Q(2) &= 4E(2) \\
Q(3) &= 0E(3) \\
Q(4) &= 4E(4)
\end{aligned}$$

which can be rewritten as a system of linear equations

$$\begin{array}{ccccccccccc}
 & & & & & & a_0 & & & = & 0 \\
a_3 & + & a_2 & + & a_1 & + & a_0 & - & b_0 & = & 1 \\
8a_3 & + & 4a_2 & + & 2a_1 & + & a_0 & - & 4b_0 & = & 8 \\
27a_3 & + & 9a_2 & + & 3a_1 & + & a_0 & & & = & 0 \\
64a_3 & + & 16a_2 & + & 4a_1 & + & a_0 & - & 4b_0 & = & 1
\end{array} .$$

(c) From the solution, we know

$$\begin{aligned}
Q(x) &= 4x^3 + x^2 + x, \\
E(x) &= x + b_0 = x.
\end{aligned}$$

Since $Q(x) = P(x)E(x)$, the recipient can compute $P(x) = Q(x)/E(x) = 4x^2 + x + 1$, the polynomial $P(x)$ from part (a) used by the sender. The error locating polynomial $E(x)$ is degree one, so there is only one error, and as $E(x) = x = x - 0$, the corrupted bit was the first one. To correct this error we evaluate $P(0) = 1$ and combine this with the two uncorrupted bits m_1, m_2, to get the original message

$$(m_0, m_1, m_2) = (1,1,4).$$

3 Berlekamp-Welch Algorithm with Fewer Errors

Note 8
Note 9

In class we derived how the Berlekamp-Welch algorithm can be used to correct k general errors, given $n+2k$ points transmitted. In real life, it is usually difficult to determine the number of errors that will occur. What if we have less than k errors? This is a follow up to the exercise posed in the notes.

Suppose Alice wants to send 1 message to Bob and wants to guard against 1 general error. She decides to encode the message with $P(x) = 4$ (on GF(7)) such that $P(0) = 4$ is the message she want to send. She then sends $P(0), P(1), P(2) = (4,4,4)$ to Bob.

(a) Suppose Bob receives the message $(4,5,4)$. Without performing Gaussian elimination explicitly, find $E(x)$ and $Q(x)$.

(b) Now, suppose there were no general errors and Bob receives the original message $(4,4,4)$. Show that the $Q(x), E(x)$ that you found in part (a) still satisfies $Q(i) = r_i E(i)$ for all $i = 0,1,2$.

(c) Verify that $E(x) = x$, $Q(x) = 4x$ is another possible set of polynomials that satisfies $Q(i) = r_i E(i)$ for all $i = 0,1,2$.

(d) Suppose you're actually trying to decode the received message $(4,4,4)$. Based on what you showed in the previous two parts, what will happen during row reduction when you try to solve for the unknowns?

(e) Prove that in general, no matter what the solution of $Q(x)$ and $E(x)$ are though, the recovered $P(x)$ will always be the same.

Solution:

(a) $E(x) = x-1$ and $Q(x) = P(x)E(x) = 4x-4$.

(b) This is true because there were no errors, so $P(i) = r_i$ for $i = 0,1,2$.

(c) Since $Q(x) = P(x)E(x)$ and $P(i) = r_i$ for $i = 0,1,2$, we must have $Q(i) = r_i E(i)$ for all $i = 0,1,2$.

(d) There are multiple solutions to the system of equations.

(e) Suppose we got two solutions $Q'(x), E'(x)$ and $Q(x), E(x)$. Since they are both solutions, by definition, we have $Q'(i) = r_i E'(i)$ and $Q(i) = r_i E(i)$ for $1 \le i \le n+2k$. Therefore, $Q'(i)E(i) = Q(i)E'(i) = r_i E(i)E'(i)$. However, $Q'(x)E(x) - Q(x)E'(x)$ is a degree $n+2k-1$ polynomial, which is 0 at $n+2k$ points. Thus, $Q'(x)E(x) = Q(x)E'(x)$ for all x, so we arrive at

$$\frac{Q'(x)}{E'(x)} = \frac{Q(x)}{E(x)}.$$

This proves that the final solution for $P(x)$ is the same.

CS 70 Spring 2024 | Discrete Mathematics and Probability Theory Seshia, Sinclair | DIS 5B

1 Countability: True or False

Note 11

(a) The set of all irrational numbers $\mathbb{R}\backslash\mathbb{Q}$ (i.e. real numbers that are not rational) is uncountable.

(b) The set of integers x that solve the equation $3x \equiv 2 \pmod{10}$ is countably infinite.

(c) The set of real solutions for the equation $x+y=1$ is countable.

For any two functions $f : Y \to Z$ and $g : X \to Y$, let their composition $f \circ g : X \to Z$ be given by $(f \circ g)(x) = f(g(x))$ for all $x \in X$. Determine if the following statements are true or false.

(d) f and g are injective (one-to-one) $\implies$ $f \circ g$ is injective (one-to-one).

(e) f is surjective (onto) $\implies$ $f \circ g$ is surjective (onto).

Solution:

(a) **True.** Proof by contradiction. Suppose the set of irrationals is countable. From Lecture note 10 we know that the set $\mathbb{Q}$ is countable. Since union of two countable sets is countable, this would imply that the set $\mathbb{R}$ is countable. But again from Lecture note 10 we know that this is not true. Contradiction!

(b) **True.** Multiplying both sides of the modular equation by 7 (the multiplicative inverse of 3 with respect to 10) we get $x \equiv 4 \pmod{10}$. The set of all intergers that solve this is $S = \{10k+4 : k \in \mathbb{Z}\}$ and it is clear that the mapping $k \in \mathbb{Z}$ to $10k+4 \in S$ is a bijection. Since the set $\mathbb{Z}$ is countably infinite, the set S is also countably infinite.

(c) **False.** Let $S \subset \mathbb{R}\times\mathbb{R}$ denote the set of all real solutions for the given equation. For any $x' \in \mathbb{R}$, the pair $(x',y') \in S$ if and only if $y' = 1-x'$. Thus $S = \{(x,1-x) : x \in \mathbb{R}\}$. Besides, the mapping x to $(x,1-x)$ is a bijection from $\mathbb{R}$ to S. Since $\mathbb{R}$ is uncountable, we have that S is uncountable too.

(d) **True.** Recall that a function $h : A \to B$ is injective iff $a_1 \neq a_2 \implies h(a_1) \neq h(a_2)$ for all $a_1, a_2 \in A$. Let $x_1, x_2 \in X$ be arbitrary such that $x_1 \neq x_2$. Since g is injective, we have $g(x_1) \neq g(x_2)$. Now, since f is injective, we have $f(g(x_1)) \neq f(g(x_2))$. Hence $f \circ g$ is injective.

(e) **False.** Recall that a function $h : A \to B$ is surjective iff $\forall b \in B, \exists a \in A$ such that $h(a) = b$. Let $g : \{0,1\} \to \{0,1\}$ be given by $g(0) = g(1) = 0$. Let $f : \{0,1\} \to \{0,1\}$ be given by $f(0) = 0$ and $f(1) = 1$. Then $f \circ g : \{0,1\} \to \{0,1\}$ is given by $(f \circ g)(0) = (f \circ g)(1) = 0$. Here f is surjective but $f \circ g$ is not surjective.

2 Counting Cartesian Products

Note 11

For two sets A and B, define the cartesian product as $A \times B = \{(a,b) : a \in A, b \in B\}$.

(a) Given two countable sets A and B, prove that $A \times B$ is countable.

(b) Given a finite number of countable sets $A_1, A_2, \dots, A_n$, prove that

$$A_1 \times A_2 \times \cdots \times A_n$$

is countable.

(c) Consider a countably infinite number of finite sets: $B_1, B_2, \dots$ for which each set has at least 2 elements. Prove that $B_1 \times B_2 \times \cdots$ is uncountable.

Solution:

(a) As shown in lecture, $\mathbb{N} \times \mathbb{N}$ is countable by creating a zigzag map that enumerates through the pairs: $(0,0), (1,0), (0,1), (2,0), (1,1), \dots$. Since A and B are both countable, there exists a bijection between each set and a subset of $\mathbb{N}$. Thus we know that $A \times B$ is countable because there is a bijection between a subset of $\mathbb{N} \times \mathbb{N}$ and $A \times B : f(i,j) = (A_i, B_j)$. We can enumerate the pairs (a,b) similarly.

(b) Proceed by induction.
Base Case: $n = 2$. We showed in part (a) that $A_1 \times A_2$ is countable since both A_1 and A_2 are countable.
Induction Hypothesis: Assume that for some $n \in \mathbb{N}$, $A_1 \times A_2 \times \cdots \times A_n$ is countable.
Induction Step: Consider $A_1 \times \cdots \times A_n \times A_{n+1}$. We know from our hypothesis that $A_1 \times \cdots \times A_n$ is countable, call it $C = A_1 \times \cdots \times A_n$. We proved in part (a) that since C is countable and A_{n+1} are countable, $C \times A_{n+1}$ is countable, which proves our claim.

(c) Let us assume that each B_i has size 2. If any of the sizes are greater than 2, that would only make the cartesian product larger. Notice that this is equivalent to the set of infinite length binary strings, which was proven to be uncountable in the notes.

Alternatively, we could provide a diagonalization argument: Assuming for the sake of contradiction that $B_1 \times B_2 \times \cdots$ is countable and its elements can be enumerated in a list:

$$\begin{gathered}
(b_{1,1}, b_{2,1}, b_{3,1}, b_{4,1}, \dots) \\
(b_{1,2}, b_{2,2}, b_{3,2}, b_{4,2}, \dots) \\
(b_{1,3}, b_{2,3}, b_{3,3}, b_{4,3}, \dots) \\
(b_{1,4}, b_{2,4}, b_{3,4}, b_{4,4}, \dots) \\
\vdots
\end{gathered}$$

where $b_{i,j}$ represents the item from set B_i that is included in the jth element of the Cartesian Product. Now consider the element $(\overline{b_{1,1}}, \overline{b_{2,2}}, \overline{b_{3,3}}, \overline{b_{4,4}}, \dots)$, where $\overline{b_{i,j}}$ represents any item from set B_i that differs from $b_{i,j}$ (i.e. any other element in the set). This is a valid element that should exist in the Cartesian Product $B_1, B_2, \ldots$, yet it is not in the enumerated list. This is a contradiction, so $B_1 \times B_2 \times \cdots$ must be uncountable.

3 Hello World!

Note 12

Determine the computability of the following tasks. If it's not computable, write a reduction or self-reference proof. If it is, write the program.

(a) You want to determine whether a program P on input x prints "Hello World!". Is there a computer program that can perform this task? Justify your answer.

(b) You want to determine whether a program P prints "Hello World!" before running the kth line in the program. Is there a computer program that can perform this task? Justify your answer.

(c) You want to determine whether a program P prints "Hello World!" in the first k steps of its execution. Is there a computer program that can perform this task? Justify your answer.

Solution:

(a) Uncomputable. We will reduce `TestHalt` to `PrintsHW`(P,x).

```
TestHalt(P, x):
  P'(x):
    run P(x) while suppressing print statements
    print("Hello World!")
  if PrintsHW(P', x):
    return true
  else:
    return false
```

If `PrintsHW` exists, `TestHalt` must also exist by this reduction. Since `TestHalt` cannot exist, `PrintsHW` cannot exist.

(b) Uncomputable. Reduce `PrintsHW`(P,x) from part (a) to this program `PrintsHWByK`(P,x,k).

```
PrintsHW(P, x):
  for i in range(len(P)):
    if PrintsHWByK(P, x, i):
      return true
  return false
```

(c) Computable. You can simply run the program until k steps are executed. If P has printed “Hello World!” by then, return true. Else, return false.

The reason that part (b) is uncomputable while part (c) is computable is that it’s not possible to determine if we ever execute a specific line because this depends on the logic of the program, but the number of computer instructions can be counted.

CS 70 Spring 2024 — Discrete Mathematics and Probability Theory — Seshia, Sinclair — DIS 6A

1 Hello World!

Note 12

Determine the computability of the following tasks. If it's not computable, write a reduction or self-reference proof. If it is, write the program.

(a) You want to determine whether a program P on input x prints "Hello World!". Is there a computer program that can perform this task? Justify your answer.

(b) You want to determine whether a program P prints "Hello World!" before running the kth line in the program. Is there a computer program that can perform this task? Justify your answer.

(c) You want to determine whether a program P prints "Hello World!" in the first k steps of its execution. Is there a computer program that can perform this task? Justify your answer.

Solution:

(a) Uncomputable. We will reduce TestHalt to PrintsHW(P,x).

```
TestHalt(P, x):
  P'(x):
    run P(x) while suppressing print statements
    print("Hello World!")
  if PrintsHW(P', x):
    return true
  else:
    return false
```

If PrintsHW exists, TestHalt must also exist by this reduction. Since TestHalt cannot exist, PrintsHW cannot exist.

(b) Uncomputable. Reduce PrintsHW(P,x) from part (a) to this program PrintsHWByK(P,x,k).

```
PrintsHW(P, x):
  for i in range(len(P)):
    if PrintsHWByK(P, x, i):
      return true
  return false
```

(c) Computable. You can simply run the program until k steps are executed. If P has printed "Hello World!" by then, return true. Else, return false.

The reason that part (b) is uncomputable while part (c) is computable is that it's not possible to determine if we ever execute a specific line because this depends on the logic of the program, but the number of computer instructions can be counted.

2 Code Reachability

Note 12

Consider triplets (M,x,L) where

- `M` is a Java program
- `x` is some input
- `L` is an integer

and the question of: if we execute $M(x)$, do we ever hit line L?

Prove this problem is undecidable.

Solution: Suppose we had a procedure that could decide the above; call it `Reachable(M, x, L)`. Consider the following example of a program deciding whether $P(x)$ halts:

```
def Halt(P, x):
    def M(t):
        run P(x)   # line 1 of M
        return     # line 2 of M
    return Reachable(M, 0, 2)
```

Program M reaches line 2 if and only if $P(x)$ halted. Thus, we have implemented a solution to the halting problem — contradiction.

3 Strings

Note 10

What is the number of strings consisting of:

(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) This is an $n+m$ length string. We choose n of those positions to be 1, and the rest will automatically be 0. Thus, the count is $\binom{n+m}{n}$. Another way of thinking about this is that there are $n+m$ positions, so we can consider $(n+m)!$ permutations. In this permutation, there are n ones, and the order of these ones doesn't actually matter. Every $n!$ way to order the ones is actually the exact same string, thus we divide by $n!$. Similarly, we divide by $m!$ to account for the zeros. Thus, we retrieve $\frac{(n+m)!}{n!m!}$.

(b) For this question, it is easier to consider the second method from the previous solution. There are $n_1+n_2+n_3$ positions, so we can consider $(n_1+n_2+n_3)!$ permutations. In this permutation, there are n_1 A's, and the order of these A's doesn't actually matter. Every $n_1!$ way to order the ones is actually the exact same string, thus we divide by $n_1!$. Similarly, we divide by $n_2!$ to account for the B's and also by $n_3!$ to account for the C's.

Alternatively, we could've used the counting positions strategy to approach this problem, though it is harder to generalize. We could consider an $n_1+n_2+n_3$ length string. First, we'll choose n_1 of those positions to be an A. Then, out of the n_2+n_3 positions left, we'll choose n_2 to be a B. Thus, the count becomes $\binom{n_1+n_2+n_3}{n_1}\binom{n_2+n_3}{n_2}$ which does evaluate to the same quantity.

(c) Using the same logic from the previous part, we generalize for a size k alphabet.

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

.

4 You'll Never Count Alone

Note 10

(a) An anagram of LIVERPOOL is any re-ordering of the letters of LIVERPOOL, i.e., any string made up of the letters L, I, V, E, R, P, O, O, L in any order. For example, IVLERPOOL and POLIVOLRE are anagrams of LIVERPOOL but PIVEOLR and CHELSEA are not. The anagram does not have to be an English word.

How many different anagrams of LIVERPOOL are there?

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

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

Solution:

(a) In this 9 letter word, the letters L and O 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 O's among themselves). Hence, there are $9!/(2!)^2$ different anagrams.

(b) $\binom{n+k}{k}$. 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").

(c) $\binom{(n-(k+1))+k}{k} = \binom{n-1}{k}$. 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.

Alternatively, we can derive a method similar to stars and bars/balls and bins; here, the restriction to positive integers means that we cannot have any empty groups. In particular, instead of arranging all of the objects (i.e. all the stars and all the bars), we can instead choose where to place the bars.

Looking at the "gaps" between the stars (i.e. the 1's), we have a total of $n-1$ places to put the bars in between the n stars. Selecting k of these positions (we can't have two bars occupy the same gap, otherwise we'd have an empty group), we have a total of $\binom{n-1}{k}$ ways to group the 1's.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 6B

1 Inclusion and Exclusion

Note 10

What is the total number of positive integers strictly less than 100 that are also coprime to 100?

Solution: It is sufficient to count the opposite: what is the total number of positive integers strictly less than 100 and *not* coprime to 100?

If a number is not coprime to 100, this means that the number is either a multiple of 2 or a multiple of 5. In this case, we have:

- 49 multiples of 2
- 19 multiples of 5
- 9 multiples of both 2 and 5

By inclusion-exclusion, the total number of positive integers not coprime to 100 is $49+19-9=59$, and there are 99 positive integers strictly less than 100.

As such, in total there are $99-59=40$ different positive integers strictly less than 100 that are coprime to 100.

2 CS70: The Musical

Note 10

Edward, one of the previous head TA's, has been hard at work on his latest project, *CS70: The Musical*. It's now time for him to select a cast, crew, and directing team to help him make his dream a reality.

(a) First, Edward would like to select directors for his musical. He has received applications from $2n$ directors. Use this to provide a combinatorial argument that proves the following identity:

$$\binom{2n}{2} = 2\binom{n}{2} + n^2.$$

(b) Edward would now like to select a crew out of n people. Use this to provide a combinatorial argument that proves the following identity: (this is called Pascal's Identity)

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

(c) There are n actors lined up outside of Edward's office, and they would like a role in the musical (including a lead role). However, he is unsure of how many individuals he would like to cast. Use this to provide a combinatorial argument that proves the following identity:

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$$

(d) Generalizing the previous part, provide a combinatorial argument that proves the following identity:

$$\sum_{k=j}^{n} \binom{n}{k}\binom{k}{j} = 2^{n-j}\binom{n}{j}.$$

Solution:

(a) Say that we would like to select 2 directors.

LHS: This is the number of ways to choose 2 directors out of the $2n$ candidates.

RHS: Split the $2n$ directors into two groups of n; one group consisting of experienced directors, or inexperienced directors (you can split arbitrarily). Then, we consider three cases: either we choose:

(a) Both directors from the group of experienced directors,

(b) Both directors from the group of inexperienced directors, or

(c) One experienced director and one inexperienced director.

The number of ways we can do each of these things is $\binom{n}{2}$, $\binom{n}{2}$, and n^2, respectively. Since these cases are mutually exclusive and cover all possibilities, it must also count the total number of ways to choose 2 directors out of the $2n$ candidates. This completes the proof.

(b) Say that we would like to select k crew members.

LHS: This is simply the number of ways to choose k crew members out of n candidates.

RHS: We select the k crew members in a different way. First, Edward looks at the first candidate he sees and decides whether or not he would like to choose the candidate. If he selects the first candidate, then Edward needs to choose $k-1$ more crew members from the remaining $n-1$ candidates. Otherwise, he needs to select all k crew members from the remaining $n-1$ candidates.

We are not double counting here - since in the first case, Edward takes the first candidate he encounters, and in the other case, we do not.

(c) In this part, Edward selects a subset of the n actors to be in his musical. Additionally, assume that he must select one individual as the lead for his musical.

LHS: Edward casts k actors in his musical, and then selects one lead among them (note that k = $\binom{k}{1}$). The summation is over all possible sizes for the cast - thus, the expression accounts for all subsets of the n actors.

RHS: From the n people, Edward selects one lead for his musical. Then, for the remaining $n-1$ actors, he decides whether or not he would like to include them in the cast. 2^{n-1} represents the amount of (possibly empty) subsets of the remaining actors. *(Note that for each actor, Edward has 2 choices: to include them, or to exclude them.)*

(d) In this part, Edward selects a subset of the n actors to be in the musical; additionally he must select j lead actors (instead of only 1 in the previous part).

LHS: Edward casts $k \geq j$ actors in his musical, then selects the j leads among them. Again, the summation is over all possible sizes for the cast (note that any cast that has $< j$ members is invalid, since Edward would be unable to select j lead actors) - thus, the expression accounts for all valid subsets of the n actors.

RHS: From the n people, Edward selects j leads for his musical. Then, for the remaining $n-j$ actors, he decides whether or not he would like to include them in the cast. Then, for the remaining $n-j$ actors, he decides whether or not he would like to include them in the cast. 2^{n-j} represents the amount of ways that Edward can do this.

3 Farmer's Market

Note 10

Suppose you want k items from the farmer's market. Count how many ways you can do this, assuming:

(a) There are pumpkins and apples at the market.

(b) There are pumpkins, apples, oranges, and pears at the market.

(c) There are n kinds of fruits at the market, and you want to end up with at least two different types of fruit.

Solution:

This is a classic "balls and bins" (also known as "stars and bars") problem.

(a) $k+1$. We can have 0 pumpkins and k apples, or 1 pumpkin and $k-1$ apples, etc. all the way to k pumpkins and 0 apples. We can equivalently think about this as k balls and 2 bins, or k stars and 1 bar, giving us $\binom{k+1}{1} = \binom{k+1}{k}$.

(b) $\binom{k+3}{3}$. We have k balls and 4 bins, or k stars and 3 bars.

(c) There are $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$ ways to choose k fruits of n types with no additional restrictions (i.e. k balls and n bins, or k stars and $n-1$ bars). n of these combinations, however, contain only one variety of fruit, so we subtract them for a total of $\binom{n+k-1}{n-1} - n = \binom{n+k-1}{k} - n$.

4 The Count

Note 10

(a) The Count is trying to choose his new 7-digit phone number. Since he is picky about his numbers, he wants it to have the property that the digits are non-increasing when read from left to right. For example, 9973220 is a valid phone number, but 9876545 is not. How many choices for a new phone number does he have?

(b) Now instead of non-increasing, they must be strictly decreasing. So 9983220 is no longer valid, while 9753210 is valid. How many choices for a new phone number does he have now?

(c) The Count now wants to make a password to secure his phone. His password must be exactly 10 digits long and can only contain the digits 0 and 1. On top of that, he also wants it to contain at least five consecutive 0's. How many possible passwords can he make?

Solution:

(a) This is actually a stars and bars problem in disguise! We have seven positions for digits, and nine dividers to partition these positions into places for nines, places for eights, etc. This is because we know that the digits are non-increasing, so all the nines (if any) must come first, then all the eights (if any), and so on. That means there are a total of 16 objects and dividers, and we are looking for where to put the nine dividers, so our answer is $\binom{16}{9}$.

(b) This can be found from just combinations. For any choice of 7 digits, there is exactly one arrangement of them that is strictly decreasing. Thus, the total number of strictly decreasing strings is exactly $\binom{10}{7}$.

(c) This problem is a bit trickier to approach, since there is a strong possibility of overcounting - it is not sufficient to just choose 5 consecutive positions to be 00000, and let the rest of the positions be arbitrary values.

One counting strategy is strategic casework - we will split up the problem into exhaustive cases based on where the run of 0's begins (we look at the leftmost zero of a run of at least 5 zeros). It can begin somewhere between the first digit and the sixth digit, inclusively.

If the run begins with the first digit, the first five digits are 0, and there are $2^5 = 32$ choices for the other 5 digits.

If the run begins after the i^{th} digit, then the $i-1^{th}$ digit must be a 1, and the other $(10-5-1=4)$ digits can be chosen arbitrarily. The other four digits can be freely chosen with $2^4 = 16$ possibilities. Thus the total number of valid passwords is $2^5 + 5 \cdot 2^4 = 112$. Note that, since there are only 10 digits, there can only be one occurence of the "100000" pattern.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 7B

1 Venn Diagram

Note 13

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?

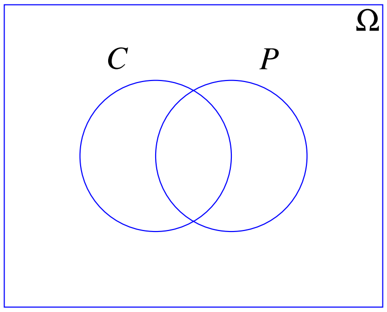
(c) What is the probability that the student works part time?

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

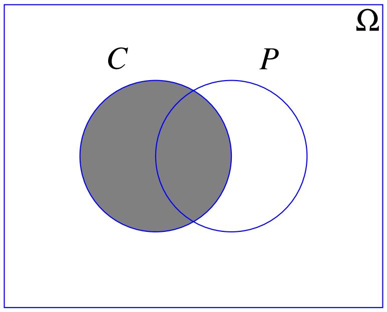
(e) What is the probability that the student belongs to a club OR works part time?

Solution:

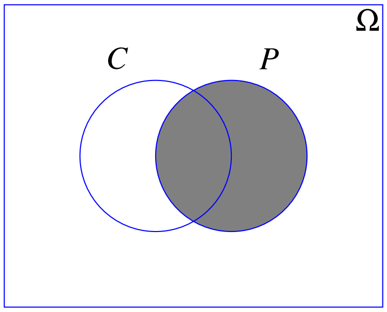
(a) The sample space will be illustrated by a Venn diagram.



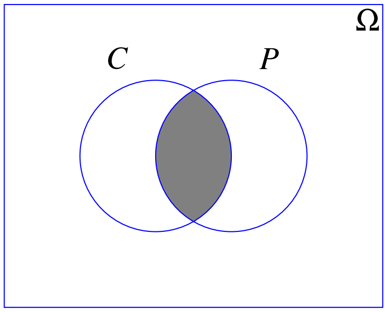
(b) $\mathbb{P}[C] = \frac{|C|}{|\Omega|} = \frac{400}{1000} = \frac{2}{5}$.



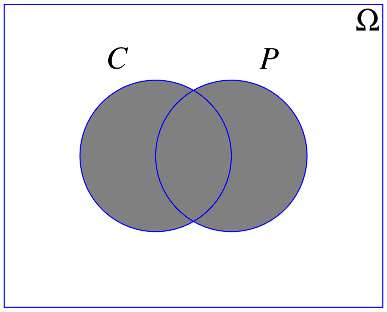
(c) $\mathbb{P}[P] = \frac{|P|}{|\Omega|} = \frac{500}{1000} = \frac{1}{2}$.



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



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



2 Flippin' Coins

Note 13

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 Sampling

Note 13

Suppose you have balls numbered $1,\dots,n$, where n is a positive integer ≥ 2, inside a coffee mug. You pick a ball uniformly at random, look at the number on the ball, replace the ball back into the coffee mug, and pick another ball uniformly at random.

(a) What is the probability that the first ball is 1 and the second ball is 2?

(b) What is the probability that the second ball's number is strictly less than the first ball's number?

(c) What is the probability that the second ball's number is exactly one greater than the first ball's number?

(d) Now, assume that after you looked at the first ball, you did *not* replace the ball in the coffee mug (instead, you threw the ball away), and then you drew a second ball as before. Now, what are the answers to the previous parts?

Solution:

(a) Out of n^2 pairs of balls that you could have chosen, only one pair $(1,2)$ corresponds to the event we are interested in, so the probability is $1/n^2$.

(b) Again, there are n^2 total outcomes. Now, we want to count the number of outcomes where the second ball's number is strictly less than the first ball's number. Similarly to the last part, we can view any outcome as an ordered pair (n_1,n_2), where n_1 is the number on the first ball, and n_2 is the number on the second ball. There are $\binom{n}{2}$ outcomes where $n_1 > n_2$; select two distinct numbers from $[1,n]$, and assign the higher number to n_1. Thus, the probability is $\frac{\binom{n}{2}}{n^2} = \frac{n-1}{2n}$.

Alternate Solution: The probability that the two balls have the same number is $n/n^2 = 1/n$, so the probability that the balls have different numbers is $1-1/n = (n-1)/n$. By symmetry, it is equally likely for the first ball to have a greater number and for the second ball to have a greater number, so we take the probability $(n-1)/n$ and divide it by two to obtain $(n-1)/(2n)$.

(c) Again, there are n^2 pairs of balls that we could have drawn, but there are $n-1$ pairs of balls which correspond to the event we are interested in: $\{(1,2),(2,3),\dots,(n-1,n)\}$. So, the probability is $(n-1)/n^2$.

(d) There are a total of $n(n-1)$ pairs of balls that we could have drawn, and only the pair $(1,2)$ corresponds to the event that we are interested in, so the probability is $1/(n(n-1))$.

The probability that the two balls are the same is now 0, but the symmetry described earlier still applies, so the probability that the second ball has a smaller number is $1/2$.

There are a total of $n(n-1)$ pairs of balls that we could have drawn, and we are interested in the $n-1$ pairs $(1,2),(2,3),\dots,(n-1,n)$ as before. Thus, the probability that the second ball is one greater than the first ball is $1/n$.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair **DIS 8A**

1 Box of Marbles

Note 14

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 Poisoned Smarties

Note 14

Supposed there are 3 people who are all owners of their own Smarties factories. Burr Kelly, being the brightest and most innovative of the owners, produces considerably more Smarties than her competitors and has a commanding 50% of the market share. Yousef See, who inherited her riches, lags behind Burr and produces 40% of the world's Smarties. Finally Stan Furd, brings up the rear with a measly 10%. However, a recent string of Smarties related food poisoning has forced the FDA investigate these factories to find the root of the problem. Through her investigations, the inspector found that 2 Smarties out of every 100 at Kelly's factory was poisonous. At See's factory, 5% of Smarties produced were poisonous. And at Furd's factory, the probability a Smarty was poisonous was 0.1.

(a) What is the probability that a randomly selected Smarty will be safe to eat?

(b) If we know that a certain Smarty didn't come from Burr Kelly's factory, what is the probability that this Smarty is poisonous?

(c) Given this information, if a randomly selected Smarty is poisonous, what is the probability it came from Stan Furd's Smarties Factory?

Solution:

(a) Let S be the event that a smarty is safe to eat. Let BK be the event that a smarty is from Burr Kelly's factory. Let YS be the event that a smarty is from Yousef See's factory. Finally, let SF be the event that a smarty is from Stan Furd's factory.

By total probability, we have

$$\begin{aligned}\mathbb{P}[S] &= \mathbb{P}[BK]\mathbb{P}[S \mid BK] + \mathbb{P}[YS]\mathbb{P}[S \mid YS] + \mathbb{P}[SF]\mathbb{P}[S \mid SF] \\ &= \frac{1}{2}\cdot\frac{49}{50} + \frac{2}{5}\cdot\frac{19}{20} + \frac{1}{10}\cdot\frac{9}{10} \\ &= \frac{49}{100} + \frac{38}{100} + \frac{9}{100} \\ &= \frac{96}{100} = \frac{24}{25} = 0.96\end{aligned}$$

Therefore the probability that a Smarty is safe to eat is 0.96.

(b) Let P be the event that a smarty is poisonous.

$$\mathbb{P}[P \mid \overline{BK}] = \frac{\mathbb{P}[\overline{BK} \cap P]}{\mathbb{P}[\overline{BK}]}$$

Since BK, YS, SF are a partition of the entire sample space, we know that if BK did not occur, then either YS occurred, or SF occurred:

$$\begin{aligned}
&= \frac{\mathbb{P}[YS \cap P]}{\mathbb{P}[\overline{BK}]} + \frac{\mathbb{P}[SF \cap P]}{\mathbb{P}[\overline{BK}]} \\
&= \frac{\mathbb{P}[P \mid YS]\mathbb{P}[YS]}{1-\mathbb{P}[BK]} + \frac{\mathbb{P}[P \mid SF]\mathbb{P}[SF]}{1-\mathbb{P}[BK]} \\
&= \frac{\frac{1}{20}\cdot\frac{2}{5}}{\frac{1}{2}} + \frac{\frac{1}{10}\cdot\frac{1}{10}}{\frac{1}{2}} = 2\cdot\frac{2}{100} + 2\cdot\frac{1}{100} \\
&= \frac{6}{100} = \frac{3}{50} = 0.06
\end{aligned}$$

(c) From Bayes' Rule, we know that:

$$\mathbb{P}[SF \mid P] = \frac{\mathbb{P}[P \mid SF]\mathbb{P}[SF]}{\mathbb{P}[P]}.$$

In part (a), we calculated the probability that any random Smarty was safe to eat; here, notice that $\mathbb{P}[P] = 1 - \mathbb{P}[S]$. This means we have

$$\begin{aligned}
\mathbb{P}[SF \mid P] &= \frac{\mathbb{P}[P \mid SF]\mathbb{P}[SF]}{1-\mathbb{P}[S]} \\
&= \frac{\frac{1}{10}\cdot\frac{1}{10}}{1-\frac{24}{25}} = \frac{\frac{1}{100}}{\frac{1}{25}} \\
&= \frac{25}{100} = \frac{1}{4} = 0.25
\end{aligned}$$

3 Pairwise Independence

Note 14

Recall that the events A_1, A_2, and A_3 are *pairwise independent* if for all $i \neq j$, A_i is independent of A_j. However, pairwise independence is a weaker statement than *mutual independence*, which requires the additional condition that $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3]$.

Suppose you roll two fair six-sided dice. Let A_1 be the event that the first die lands on 1, let A_2 be the event that the second die lands on 6, and let A_3 be the event that the two dice sum to 7.

(a) Compute $\mathbb{P}[A_1]$, $\mathbb{P}[A_2]$, and $\mathbb{P}[A_3]$.

(b) Are A_1 and A_2 independent?

(c) Are A_2 and A_3 independent?

(d) Are A_1, A_2, and A_3 pairwise independent?

(e) Are A_1, A_2, and A_3 mutually independent?

Solution:

(a) We have that $\mathbb{P}[A_1] = \mathbb{P}[A_2] = \frac{1}{6}$, since we have a $\frac{1}{6}$ probability of getting a particular number on a fair die.

Since there are 6 ways in which the two dice can sum to 7 (i.e. $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$), we have $\mathbb{P}[A_3] = \frac{1}{6}$ as well.

(b) We want to determine whether $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2]$. We already found the probabilities of A_1 and A_2 from part (a), so let's look at $\mathbb{P}[A_1 \cap A_2]$. There's only one possible outcome where the first die is a 1 and the second die is a 6, so this gives a probability of $\mathbb{P}[A_1 \cap A_2] = \frac{1}{36}$.

Since $\mathbb{P}[A_1]\mathbb{P}[A_2] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_1 \cap A_2]$, these two events are independent.

(c) We want to determine whether $\mathbb{P}[A_2 \cap A_3] = \mathbb{P}[A_2]\mathbb{P}[A_3]$. We already found the probabilities of A_2 and A_3 from part (a), so let's look at $\mathbb{P}[A_2 \cap A_3]$. These two events both occur if the second die lands on a 6, and the two dice sum to 7. There's only one way that this can happen, i.e. the first die must be a 1, so the intersection has probability $\mathbb{P}[A_2 \cap A_3] = \frac{1}{36}$.

Since $\mathbb{P}[A_2]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_2 \cap A_3]$, these two events are independent.

(d) To see whether the three events are pairwise independent, we need to ensure that all pairs of events are independent. We've already checked that A_1 and A_2 are independent, and that A_2 and A_3 are independent, so it suffices to check whether A_1 and A_3 are independent.

Similar to the previous two parts, the intersection $A_1 \cap A_3$ means that the first die must land on a 1, and the two dice sum to 7. There's only one way for this to happen, i.e. the second die must land on a 6, so the probability is $\mathbb{P}[A_1 \cap A_3] = \frac{1}{36}$.

Since $\mathbb{P}[A_1]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_1 \cap A_3]$, these two events are also independent. Since we've now shown that all possible pairs of events are independent, A_1, A_2, and A_3 are indeed pairwise independent.

(e) Mutual independence requires the additional constraint that $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3]$. We've found the individual probabilities of these events in part (a), so we only need to compute $\mathbb{P}[A_1 \cap A_2 \cap A_3]$.

Here, we must have that the first die lands on 1, the second die lands on 6, and the sum of the two dice is equal to 7. There's only one way for this to happen, i.e. the first die is a 1 and the second die is a 6, so the probability of the intersection of all three events is $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \frac{1}{36}$.

However, since $\mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} \neq \frac{1}{36} = \mathbb{P}[A_1 \cap A_2 \cap A_3]$, these three events are not mutually independent.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 8B

1 Probability Potpourri

Note 13
Note 14

Provide brief justification for each part.

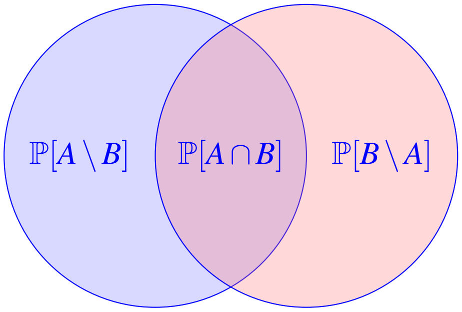
(a) For two events A and B in any probability space, show that $\mathbb{P}[A \setminus B] \geq \mathbb{P}[A] - \mathbb{P}[B]$.

(b) 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.

(c) If A and B are disjoint, does that imply they're independent?

Solution:

(a) It can be helpful to first draw out a Venn diagram:



We can see here that $\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]$, and that $\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[B \setminus A]$. Looking at the RHS, we have

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

(b) Using the total probability rule, we have

$$\mathbb{P}[D] = \mathbb{P}[D \cap C] + \mathbb{P}[D \cap \overline{C}] = \mathbb{P}[D \mid C] \cdot \mathbb{P}[C] + \mathbb{P}[D \mid \overline{C}] \cdot \mathbb{P}[\overline{C}].$$

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.

(c) No; if two events are disjoint, we cannot conclude they are independent. Consider a roll of a fair six-sided die. Let A be the event that we roll a 1, and let B be the event that we roll a 2. Certainly A and B are disjoint, as $\mathbb{P}[A \cap B] = 0$. But these events are not independent: $\mathbb{P}[B \mid A] = 0$, but $\mathbb{P}[B] = 1/6$.

Since disjoint events have $\mathbb{P}[A \cap B] = 0$, we can see that the only time when disjoint A and B are independent is when either $\mathbb{P}[A] = 0$ or $\mathbb{P}[B] = 0$.

2 Easter Eggs

Note 14

You made the trek to Soda for a Spring Break-themed homework party, and every attendee gets to leave with a party favor. You're given a bag with 20 chocolate eggs and 40 (empty) plastic eggs. You pick 5 eggs (uniformly) without replacement.

(a) What is the probability that the first egg you drew was a chocolate egg?

(b) What is the probability that the second egg you drew was a chocolate egg?

(c) Given that the first egg you drew was an empty plastic one, what is the probability that the fifth egg you drew was also an empty plastic egg?

Solution:

(a) $\mathbb{P}[\text{chocolate egg}] = \dfrac{20}{60} = \dfrac{1}{3}$.

(b) Long calculation using Total Probability Rule: let C_i denote the event that the ith egg is chocolate, and P_i denote the event that the ith egg is plastic. We have

$$\begin{aligned}\mathbb{P}[C_2] &= \mathbb{P}[C_1 \cap C_2] + \mathbb{P}[P_1 \cap C_2] \\ &= \mathbb{P}[C_1]\mathbb{P}[C_2 \mid C_1] + \mathbb{P}[P_1]\mathbb{P}[C_2 \mid P_1] \\ &= \frac{1}{3} \cdot \frac{19}{59} + \frac{2}{3} \cdot \frac{20}{59} \\ &= \frac{1}{3}.\end{aligned}$$

Short calculation: By symmetry, this is the same probability as part (a), $1/3$. This is because we don't know what type of egg was picked on the first draw, so the distribution for the second egg is the same as that of the first. To see this rigorously observe that $\mathbb{P}[C_2 \cap P_1] = \mathbb{P}[P_2 \cap C_1]$ and, thus:

$$\begin{aligned}\mathbb{P}[C_2] &= \mathbb{P}[C_2 \cap C_1] + \mathbb{P}[C_2 \cap P_1] \\ &= \mathbb{P}[C_2 \cap C_1] + \mathbb{P}[P_2 \cap C_1] \\ &= \mathbb{P}[C_1]\end{aligned}$$

(c) By symmetry, since we don't know any information about the 2nd, 3rd, or 4th eggs, we have

$$\mathbb{P}[\text{5th egg = plastic} \mid \text{1st egg = plastic}] = \mathbb{P}[\text{2nd egg = plastic} \mid \text{1st egg = plastic}] = \frac{39}{59}.$$

Rigorously, notice that $\mathbb{P}[C_5 \cap P_2 \mid P_1] = \mathbb{P}[P_5 \cap C_2 \mid P_1]$ and therefore:

$$\begin{aligned}\mathbb{P}[P_5 \mid P_1] &= \mathbb{P}[P_5 \cap C_2 \mid P_1] + \mathbb{P}[P_5 \cap P_2 \mid P_1] \\ &= \mathbb{P}[C_5 \cap P_2 \mid P_1] + \mathbb{P}[P_5 \cap P_2 \mid P_1] \\ &= \mathbb{P}[P_2 \mid P_1]\end{aligned}$$

One could also brute force this with Total Probability Rule (like in the previous part), but the calculation is quite tedious.

3 Balls and Bins

Note 14

Suppose you 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. Let m be the number of subsets of k bins out of the total n bins. If we assume A_i is the event that the ith set of k bins is empty. Then we can write A as the union of A_i's:

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

Compute m in terms of n and k, and use the union bound to give an upper bound on the probability $\mathbb{P}[A]$.

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

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

(f) 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) Note that this is a uniform sample space, with outcomes representing all possible ways to throw each ball individually into the bins. Here, $|\Omega| = n^n$, as each of the n balls has n possible bins to fall into, and out of these possibilities, $(n-1)^n$ of them leave the first bin empty—each ball would then have $n-1$ possible bins to fall into. This gives us an overall probability $\left(\frac{n-1}{n}\right)^n$ that the first bin is empty.

Equivalently, we can note that each throw is independent of all of the other throws. Since 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) Similar to the previous part, we have the same uniform sample space of size n^n. Now, there are a total of $(n-k)^n$ possible ways to throw the balls into bins such that the first k bins are empty—each ball has $n-k$ possible bins to fall into.

Alternatively, we can similarly make use of independence. Since 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

$$\mathbb{P}[A] = \mathbb{P}\left[\bigcup_{i=1}^{m} A_i\right] \leq \sum_{i=1}^{m} \mathbb{P}[A_i].$$

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,

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

(d) Using Bayes' Rule:

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

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 $\frac{n-2}{n-1}$. For n total balls, this probability is $\left(\frac{n-2}{n-1}\right)^n$.

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

(f) In part (c) we calculated the probability that the second bin is empty given that the first bin is empty: $\left(\frac{n-2}{n-1}\right)^n$. The probability that the second bin is empty (without any prior information) is $\left(\frac{n-1}{n}\right)^n$. Since these probabilities are not equal, the events are dependent.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 9A

1 Head Count

Note 15

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) What is $\mathbb{P}[X = k]$, for some $0 \leq k \leq 20$?

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

(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) There are a total of $\binom{20}{k}$ ways to select k coins to be heads. The probability that the selected k coins to be heads is $(\frac{2}{5})^k$, and the probability that the rest are tails is $(\frac{3}{5})^{20-k}$. Putting this together, we have

$$\mathbb{P}[X = k] = \binom{20}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{20-k}.$$

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

(c)

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

(d)

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

2 Family Planning

Note 15

Mr. and Mrs. Johnson decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Johnsons have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C = 1$	$C = 2$	$C = 3$
$G = 0$			
$G = 1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}[G = 0]$	
$\mathbb{P}[G = 1]$	

$\mathbb{P}[C = 1]$	$\mathbb{P}[C = 2]$	$\mathbb{P}[C = 3]$

(d) Are G and C independent?

(e) What is the expected number of girls the Johnsons will have? What is the expected number of children that the Johnsons will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Johnsons can have: $\Omega = \{g, bg, bbg, bbb\}$. The probabilities of these sample points are:

$$\mathbb{P}[g] = \frac{1}{2}$$
$$\mathbb{P}[bg] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$\mathbb{P}[bbg] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$
$$\mathbb{P}[bbb] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

(b)

	$C = 1$	$C = 2$	$C = 3$
$G = 0$	0	0	$\mathbb{P}[bbb] = 1/8$
$G = 1$	$\mathbb{P}[g] = 1/2$	$\mathbb{P}[bg] = 1/4$	$\mathbb{P}[bbg] = 1/8$

(c) Marginal distribution for G:

$$\mathbb{P}[G=0] = 0+0+\frac{1}{8} = \frac{1}{8}$$
$$\mathbb{P}[G=1] = \frac{1}{2}+\frac{1}{4}+\frac{1}{8} = \frac{7}{8}$$

Marginal distribution for C:

$$\mathbb{P}[C=1] = 0+\frac{1}{2} = \frac{1}{2}$$
$$\mathbb{P}[C=2] = 0+\frac{1}{4} = \frac{1}{4}$$
$$\mathbb{P}[C=3] = \frac{1}{8}+\frac{1}{8} = \frac{1}{4}$$

(d) No, G and C are not independent. If two random variables are independent, then

$$\mathbb{P}[X=x, Y=y] = \mathbb{P}[X=x]\mathbb{P}[Y=y].$$

To show this dependence, consider an entry in the joint distribution table, such as $\mathbb{P}[G=0, C=3] = 1/8$. This is not equal to $\mathbb{P}[G=0]\mathbb{P}[C=3] = (1/8)\cdot(1/4) = 1/32$, so the random variables are not independent.

(e) We can apply the definition of expectation directly for this problem, since we've computed the marginal distribution for both random variables.

$$\mathbb{E}[G] = 0\cdot\mathbb{P}[G=0]+1\cdot\mathbb{P}[G=1] = 1\cdot\frac{7}{8} = \frac{7}{8}$$
$$\mathbb{E}[C] = 1\cdot\mathbb{P}[C=1]+2\cdot\mathbb{P}[C=2]+3\cdot\mathbb{P}[C=3] = 1\cdot\frac{1}{2}+2\cdot\frac{1}{4}+3\cdot\frac{1}{4} = \frac{7}{4}$$

3 Pullout Balls

Note 15

Suppose you have a bag containing four balls numbered $1, 2, 3, 4$.

(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

$$\mathbb{E}[X] = \sum_x x\cdot\mathbb{P}[X=x] = 1\times\frac{1}{4}+2\times\frac{1}{4}+3\times\frac{1}{4}+4\times\frac{1}{4} = 2.5.$$

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 $X(\omega) = 2.5$).

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

$$\mathbb{E}[Y] = \frac{1}{\binom{4}{2}}(1\cdot 2 + 1\cdot 3 + 1\cdot 4 + 2\cdot 3 + 2\cdot 4 + 3\cdot 4) = \frac{2+3+4+6+8+12}{6} = \frac{35}{6}.$$

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair DIS 9B

1 Linearity

Note 15

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? (*Hint*: Consider where the sequence "book" can appear in the string.)

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

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

Then the expected total number of tickets you receive, by linearity of expectation, is

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

Note that $10\left(3 \cdot \frac{1}{3}\right)$ and $20\left(4 \cdot \frac{1}{5}\right)$ matches the expression directly gotten using the expected value of a binomial random variable.

(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 Head Count II

Note 19

Consider a coin with $\mathbb{P}[\text{Heads}] = 3/4$. Suppose you flip the coin until you see heads for the first time, and define X to be the number of times you flipped the coin.

(a) What is $\mathbb{P}[X = k]$, for some $k \geq 1$?

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

(c) What is $\mathbb{P}[X > k]$, for some $k \geq 0$?

(d) What is $\mathbb{P}[X < k]$, for some $k \geq 1$?

(e) What is $\mathbb{P}[X > k \mid X > m]$, for some $k \geq m \geq 0$? How does this relate to $\mathbb{P}[X > k - m]$?

(f) Suppose $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(q)$ are independent. Find the distribution of $\min(X, Y)$ and justify your answer.

Solution:

(a) If we flipped k times, then we had $k-1$ tails and 1 head, in that order, giving us

$$\mathbb{P}[X = k] = \frac{3}{4}\left(1 - \frac{3}{4}\right)^{k-1} = \frac{3}{4}\left(\frac{1}{4}\right)^{k-1}.$$

(b) $X \sim \text{Geometric}(\frac{3}{4})$

(c) If we had to flip *more than* k times before seeing our first heads, then our first k flips must have been tails, giving us

$$\mathbb{P}[X > k] = \left(1 - \frac{3}{4}\right)^{k} = \left(\frac{1}{4}\right)^{k}.$$

You can alternatively write as the sum $\sum_{i=k+1}^{\infty} \mathbb{P}[X = i] = \sum_{i=k+1}^{\infty} \frac{3}{4} * (\frac{1}{4})^{i-1} = \frac{3}{4} * (\frac{1}{4})^{k} * \frac{1}{1-1/4} = (\frac{1}{4})^{k}$ using the formula for an infinite geometric sum

(d) Notice $\mathbb{P}[X < k] = 1 - \mathbb{P}[X \geq k] = 1 - \mathbb{P}[X > k-1]$ since X can only take on integer values. Along similar lines to the previous part, we then have

$$\mathbb{P}[X < k] = 1 - \mathbb{P}[X > k-1] = 1 - \left(1 - \frac{3}{4}\right)^{k-1} = 1 - \left(\frac{1}{4}\right)^{k-1}.$$

(e) By part (c), we have

$$\mathbb{P}[X > k \mid X > m] = \frac{\mathbb{P}[X > k \cap X > m]}{\mathbb{P}[X > m]} = \frac{\mathbb{P}[X > k]}{\mathbb{P}[X > m]} = \left(\frac{1}{4}\right)^{k-m}.$$

However, note that this is exactly $\mathbb{P}[X > k-m]$. The reason this makes sense is that if we want to compute the probability that the first heads occurs after k flips, and we know that the first heads occurs after m flips, then the first m flips are tails. Thus, by the independence of the coin flips, the first m flips don't matter, and so we only need to compute the probability that the first heads occurs after $k-m$ flips. This is called the **memorylessness property** of the geometric distribution.

(f) Let X be the number of coins we flip until we see a heads from flipping a coin with bias p, and let Y similarly be the number of coins we flip until we see a heads from flipping a coin with bias q.

Imagine we flip the bias p coin and the bias q coin at the same time. The minimum of the two random variables represents how many simultaneous flips occur before at least one head is seen.

The probability of not seeing a head at all on any given simultaneous flip is $(1-p)(1-q)$; this corresponds to a failure. This means that the probability that there will be a success on any particular trial is $1-(1-p)(1-q) = p+q-pq$. Therefore, $\min(X,Y) \sim \text{Geometric}(p+q-pq)$.

Alternative 1: We can also solve this algebraically. The probability that $\min(X,Y) = k$ for some positive integer k is the probability that the first $k-1$ coin flips for both X and Y were tails, and we get heads on the kth toss (this can come from either X or Y). Specifically, this occurs with probability

$$((1-p)(1-q))^{k-1} \cdot (p+q-pq)$$

We recognize this as the formula for a geometric random variable with parameter $p+q-pq$.

Alternative 2: An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. Let $Z = \min(X,Y)$. We can work with $\mathbb{P}[Z \geq k]$ rather than with $\mathbb{P}[Z = k]$; clearly the values $\mathbb{P}[Z \geq k]$ specify the values $\mathbb{P}[Z = k]$ since $\mathbb{P}[Z = k] = \mathbb{P}[Z \geq k] - \mathbb{P}[Z \geq (k+1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned}
\mathbb{P}[Z \geq k] &= \mathbb{P}[\min(X,Y) \geq k] \\
&= \mathbb{P}[(X \geq k) \cap (Y \geq k)] \\
&= \mathbb{P}[X \geq k] \cdot \mathbb{P}[Y \geq k] && \text{since } X,Y \text{ are independent} \\
&= (1-p)^{k-1}(1-q)^{k-1} && \text{since } x,Y \text{ are geometric} \\
&= ((1-p)(1-q))^{k-1} \\
&= (1-p-q+pq)^{k-1}.
\end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p+q-pq$, thus we can conclude that $Z \sim \mathit{Geom}(p+q-pq)$, which is the same result as before!

3 Shuttles and Taxis at Airport

Note 19

In front of terminal 3 at San Francisco Airport is a pickup area where shuttles and taxis arrive according to a Poisson distribution. The shuttles arrive at a rate $\lambda_1 = 1/20$ (i.e. 1 shuttle per 20 minutes) and the taxis arrive at a rate $\lambda_2 = 1/10$ (i.e. 1 taxi per 10 minutes) starting at 00:00. The shuttles and the taxis arrive independently.

(a) What is the distribution of the following:

(i) The number of taxis that arrive between times 00:00 and 00:20?

(ii) The number of shuttles that arrive between times 00:00 and 00:20?

(iii) The total number of pickup vehicles that arrive between times 00:00 and 00:20?

(b) What is the probability that exactly 1 shuttle and 3 taxis arrive between times 00:00 and 00:20?

(c) Given that exactly 1 pickup vehicle arrived between times 00:00 and 00:20, what is the conditional probability that this vehicle was a taxi?

(d) Suppose you reach the pickup area at 00:20. You learn that you missed 3 taxis and 1 shuttle in those 20 minutes. What is the probability that you need to wait for more than 10 mins until either a shuttle or a taxi arrives?

Solution:

(a) (i) Let $T([0,20])$ denote the number of taxis that arrive between times 00:00 and 00:20. This interval has length 20 minutes, so the number of taxis $T([0,20])$ arriving in this interval is distributed according to $\text{Poisson}(\lambda_2 \cdot 20) = \text{Poisson}(2)$, i.e.

$$\mathbb{P}[T([0,20]) = t] = \frac{2^t e^{-2}}{t!}, \text{ for } t = 0,1,2,\ldots.$$

(ii) Let $S([0,20])$ denote the number of shuttles that arrive between times 00:00 and 00:20. This interval has length 20 minutes, so the number of shuttles $S([0,20])$ arriving in this interval is distributed according to $\text{Poisson}(\lambda_1 \cdot 20) = \text{Poisson}(1)$, i.e.

$$\mathbb{P}[S([0,20]) = s] = \frac{1^s e^{-1}}{s!}, \text{ for } s = 0,1,2,\ldots.$$

(iii) Let $N([0,20]) = S([0,20]) + T([0,20])$ denote the total number of pickup vehicles (taxis and shuttles) arriving between times 00:00 and 00:20. Since the sum of independent Poisson random variables is Poisson distributed with parameter given by the sum of the individual parameters, we have $N[(0,20)] \sim \text{Poisson}(3)$, i.e.

$$\mathbb{P}[N([0,20]) = n] = \frac{3^n e^{-3}}{n!}, \text{ for } n = 0,1,2,\ldots.$$

(b) We have

$$\mathbb{P}[T([0,20]) = 3] = \frac{2^3 e^{-2}}{3!} \text{ and } \mathbb{P}[S([0,20]) = 1] = \frac{1^1 e^{-1}}{1!}.$$

Since the taxis and the shuttles arrive independently, the probability that exactly 3 taxis and 1 shuttle arrive in this interval is given by the product of their individual probabilities, i.e.

$$\frac{2^3 e^{-2}}{3!} \frac{1^1 e^{-1}}{1!} = \frac{4}{3} e^{-3} \approx 0.0664.$$

(c) Let A be the event that exactly 1 taxi arrives between times 00:00 and 00:20. Let B be the event that exactly 1 vehicle arrives between times 00:00 and 00:20. We have

$$\mathbb{P}[B] = \frac{3^1 e^{-3}}{1!}.$$

Event $A \cap B$ is the event that exactly 1 taxi and 0 shuttles arrive between times 00:00 and 00:20. Hence

$$\mathbb{P}[A \cap B] = \frac{2^1 e^{-2}}{1!} \frac{1^0 e^{-1}}{0!}.$$

Thus, we get

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = 2/3.$$

(d) The event that you need to wait for more than 10 minutes starting 00:20 is equivalent to the event that no vehicle arrives between times 00:20 and 00:30. Let $N[20,30]$ denote the number of vehicles that arrive between times 00:20 and 00:30. This interval has length 10 minutes, so $N[(20,30)] \sim \text{Poisson}((\lambda_1 + \lambda_2) \cdot 10) = \text{Poisson}(3/2)$. Since Poisson arrivals in disjoint intervals are independent, we have

$$\mathbb{P}\left[N([20,30]) = 0 \,|\, T([0,20]) = 3, S([0,20]) = 1\right] = \mathbb{P}[N([20,30]) = 0] \sim \frac{1.5^0 e^{-1.5}}{0!} = e^{-1.5} \approx 0.2231.$$

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair
DIS 11A

1 Student Life

Note 19

In an attempt to avoid having to do laundry often, Marcus comes up with a system. Every night, he designates one of his shirts as his dirtiest shirt. In the morning, he randomly picks one of his shirts to wear. If he picked the dirtiest one, he puts it in a dirty pile at the end of the day (a shirt in the dirty pile is not used again until it is cleaned).

When Marcus puts his last shirt into the dirty pile, he finally does his laundry, and again designates one of his shirts as his dirtiest shirt (laundry isn't perfect) before going to bed. This process then repeats.

(a) If Marcus has n shirts, what is the expected number of days that transpire between laundry events? Your answer should be a function of n involving no summations.

(b) Say he gets even lazier, and instead of organizing his shirts in his dresser every night, he throws his shirts randomly onto one of n different locations in his room (one shirt per location), designates one of his shirts as his dirtiest shirt, and one location as the dirtiest location.

In the morning, if he happens to pick the dirtiest shirt, *and* the dirtiest shirt was in the dirtiest location, then he puts the shirt into the dirty pile at the end of the day and does not throw any future shirts into that location and also does not consider it as a candidate for future dirtiest locations (it is too dirty).

What is the expected number of days that transpire between laundry events now? Again, your answer should be a function of n involving no summations.

Solution:

(a) The number of days that it takes for him to throw a shirt into the dirty pile can be represented as a geometric RV. For the first shirt, this is the geometric RV with $p = 1/n$. We can see this by noticing that every day up to the day he picks the dirtiest shirt, the probability of getting the dirtiest shirt remains $1/n$.

We'll call X_i the number of days that go until he throws the ith shirt into the dirty pile. Since on the ith shirt, there are $n-i+1$ shirts left, we get that $X_i \sim \text{Geometric}(1/(n-i+1))$. The number of days until he does his laundry is a sum of these variables. Therefore, we can get the following result:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1) = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

(b) For this part we can use a similar approach but the probability for X_i becomes $1/(n-i+1)^2$. This is because the dirtiest shirt falls into the dirtiest spot with probability $1/(n-i+1)$ and we pick it after that with probability $1/(n-i+1)$, so the probability of picking the dirtiest shirt from the dirtiest spot for the ith shirt is $1/(n-i+1)^2$. Using the same approach, we get the following sum:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1)^2 = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

2 Elevator Variance

Note 16

A building has n upper floors numbered $1, 2, \dots, n$, 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 $I_1, \dots, I_n$, 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 $\operatorname{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2$ 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

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

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

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

3 Covariance

Note 16

(a) We have a bag of 5 red and 5 blue balls. We take two balls uniformly at random from the bag without replacement. Let X_1 and X_2 be indicator random variables for the events of the first and second ball being red, respectively. What is $\text{cov}(X_1,X_2)$? Recall that $\text{cov}(X,Y) = \mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y]$.

(b) Now, we have two bags A and B, with 5 red and 5 blue balls each. Draw a ball uniformly at random from A, record its color, and then place it in B. Then draw a ball uniformly at random from B and record its color. Let X_1 and X_2 be indicator random variables for the events of the first and second draws being red, respectively. What is $\text{cov}(X_1,X_2)$?

Solution:

(a) We can use the formula $\text{cov}(X_1,X_2) = \mathbb{E}[X_1X_2]-\mathbb{E}[X_1]\,\mathbb{E}[X_2]$.

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{5}{10}\times 1+\frac{5}{10}\times 0 = \frac{1}{2},\\
\mathbb{E}[X_2] &= \frac{5}{10}\times 1+\frac{5}{10}\times 0 = \frac{1}{2},\\
\mathbb{E}[X_1X_2] &= \frac{5}{10}\cdot\frac{4}{9}\times 1+\left(1-\frac{5}{10}\cdot\frac{4}{9}\right)\times 0 = \frac{2}{9}.
\end{aligned}$$

Therefore,

$$\text{cov}(X_1,X_2) = \mathbb{E}[X_1X_2]-\mathbb{E}[X_1]\,\mathbb{E}[X_2] = \frac{2}{9}-\frac{1}{2}\times\frac{1}{2} = -\frac{1}{36}.$$

(b) Again, we use the formula $\text{cov}(X_1,X_2) = \mathbb{E}[X_1X_2]-\mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{5}{10}\times 1+\frac{5}{10}\times 0 = \frac{1}{2}\\
\mathbb{E}[X_2] &= \left(\frac{5}{10}\times\frac{6}{11}+\frac{5}{10}\times\frac{5}{11}\right)\times 1+\left(\frac{5}{10}\times\frac{5}{11}+\frac{5}{10}\times\frac{6}{11}\right)\times 0 = \frac{1}{2}\\
\mathbb{E}[X_1X_2] &= \frac{5}{10}\times\frac{6}{11}\times 1 = \frac{30}{110}.
\end{aligned}$$

Therefore,

$$\mathbb{E}[X_1X_2]-\mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{30}{110}-\frac{1}{4} = \frac{1}{44}.$$

Note that in part (a), if one event happened, the other would be less likely to happen, and thus the covariance was negative. Similarly, in part (b), if one event happened, the other would be more likely to happen, and thus the covariance was positive.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 11B

1 Probabilistic Bounds

Note 17

A random variable X has variance $\operatorname{Var}(X) = 9$ and expectation $\mathbb{E}[X] = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

(a) $\mathbb{E}[X^2] = 13$.

(b) $\mathbb{P}[X = 2] > 0$.

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

(d) $\mathbb{P}[X \leq 1] \leq 8/9$.

(e) $\mathbb{P}[X \geq 6] \leq 9/16$.

Solution:

(a) TRUE. Since $9 = \operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2^2$, we have $\mathbb{E}[X^2] = 9 + 4 = 13$.

(b) FALSE. It is not necessary for a random variable to be able to take on its mean as a value. As one possible counterexample, construct a random variable X that satisfies the conditions in the question but does not take on the value 2.

A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 1/2$, and $a \neq b$ with both $a, b \leq 10$. The expectation must be 2, so we have $a/2 + b/2 = 2$. The variance is 9, so $\mathbb{E}[X^2] = 13$ (from Part (a)) and $a^2/2 + b^2/2 = 13$. Solving for a and b, we get $\mathbb{P}[X = -1] = \mathbb{P}[X = 5] = 1/2$ as a counterexample.

(c) FALSE. The median of a random variable is not necessarily the mean, unless it is symmetric. As one possible counterexample, construct a random variable X that satisfies the conditions in the question but does not have an equal chance of being less than or greater than 2.

A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = p, \mathbb{P}[X = b] = 1 - p$. Here, we use the same approach as part (b) except with a generic p, since we want $p \neq 1/2$. The expectation must be 2, so we have $pa + (1 - p)b = 2$. The variance is 9, so $\mathbb{E}\left[X^2\right] = 13$ and $pa^2 + (1 - p)b^2 = 13$. Solving for a and b, we find the relation $b = 2 \pm 3/\sqrt{x}$, where $x = (1 - p)/p$. Then, we can find an example by plugging in values for x so that $a, b \leq 10$ and $p \neq 1/2$. One such counterexample is $\mathbb{P}[X = -7] = 1/10, \mathbb{P}[X = 3] = 9/10$.

(d) TRUE. Let $Y = 10 - X$. Since X is never exceeds 10, Y is a non-negative random variable. By Markov's inequality,

$$\mathbb{P}[10 - X \geq a] = \mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a} = \frac{\mathbb{E}[10 - X]}{a} = \frac{8}{a}.$$

Setting $a = 9$, we get $\mathbb{P}[X \leq 1] = \mathbb{P}[10 - X \geq 9] \leq 8/9$.

As a side note, if we were to try Chebyshev's inequality instead, noting that

$$\mathbb{P}[X \leq 1] + \mathbb{P}[X \geq 3] = \mathbb{P}[|X - 2| \geq 1] = \mathbb{P}[|X - \mathbb{E}[X]| \geq 1],$$

we'd get

$$\mathbb{P}[X \leq 1] \leq \mathbb{P}[X \leq 1] + \mathbb{P}[X \geq 3] = \mathbb{P}[|X - 2| \geq 1] \leq \frac{\mathrm{Var}(X)}{1} = 9,$$

which is an unhelpful bound.

(e) TRUE. Chebyshev's inequality says $\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \mathrm{Var}(X)/a^2$. If we set $a = 4$, we have

$$\mathbb{P}[|X - 2| \geq 4] \leq \frac{9}{16}.$$

Now we observe that $\mathbb{P}[X \geq 6] \leq \mathbb{P}[|X - 2| \geq 4]$, because the event $X \geq 6$ is a subset of the event $|X - 2| \geq 4$.

As a side note, we can't apply Markov's inequality here; as-is, X is not nonnegative, and if we did the same transformation $Y = 10 - X$ from before, we'd want an upper bound on $\mathbb{P}[10 - X \leq 10 - 6] = \mathbb{P}[Y \leq 4]$, which we cannot do with Markov's inequality; it only gives an upper bound on probabilities of the form $\mathbb{P}[Y \geq a]$.

2 Vegas

Note 17

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 their coin. Assume that each person is independently likely to carry a fair or a trick coin.

(a) Let X be the proportion of coin flips which are heads. Find $\mathbb{E}[X]$.

(b) Given the results of your experiment, how should you estimate p? (*Hint:* Construct an unbiased estimator for p using part (a). Recall that $\hat{p}$ is an unbiased estimator if $\mathbb{E}[\hat{p}] = p$.)

(c) How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

Solution:

(a) Let X_i be the indicator that the ith person's coin flips heads. Then $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. Applying linearity, we have

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_i].$$

By total probability,

$$\mathbb{E}[X_i] = p \cdot 1 + (1-p) \cdot \frac{1}{2} = \frac{1}{2}(p+1).$$

(b) 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. In other words, we measure X, the fraction of people whose coin flips heads. How can we use this observation to construct $\hat{p}$? From (a), $\mathbb{E}[X] = \frac{1}{2}(p+1)$. By applying (reverse) linearity to isolate p, we find that

$$p = 2\,\mathbb{E}[X] - 1 = \mathbb{E}[2X-1].$$

Thus, our estimator $\hat{p}$ should be $2X-1$.

(c) We want to find n such that $\mathbb{P}[|\hat{p}-p| \le 0.05] > 0.95$. 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:

$$\mathbb{P}[|\hat{p}-p| > 0.05] \le \mathbb{P}[\hat{p}-p| \ge 0.05] \le \frac{\operatorname{Var}[\hat{p}]}{0.05^2}$$

If $\frac{\operatorname{Var}(\hat{p})}{0.05^2} \le 0.05$, then we have $\mathbb{P}[|\hat{p}-p| > 0.05] \le 0.05$ as desired. So, we want n such that $\operatorname{Var}(\hat{p}) \le 0.05^3$.

$$\operatorname{Var}(\hat{p}) = \operatorname{Var}(2X-1) = 4\operatorname{Var}(X) = \frac{4}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\operatorname{Var}(X_1).$$

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

$$\operatorname{Var}[\hat{p}] \le \frac{4}{n}\frac{1}{4} = \frac{1}{n}.$$

So, we choose n such that $\frac{1}{n} \le 0.05^3$, giving $n \ge \frac{1}{0.05^3} = 8000$.

3 Working with the Law of Large Numbers

Note 17

(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. Intuitively, the more tosses we have, the harder it gets for *exactly* half of the tosses to be heads; more tosses gives more of a restriction. In extremes, compare the probability of getting exactly 1 head out of 2 tosses (this is 0.5), and the probability of getting *exactly* 500,000 heads out of a million tosses; the latter is much much smaller than 0.5, because we're targeting such a specific number.

More rigorously, we can 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
Spring 2024 Seshia, Sinclair **DIS 12A**

1 Continuous Intro

Note 21

(a) Is

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

a valid density function? Why or why not? Is it a valid CDF? Why or why not?

(b) Calculate the PDF $f_X(x)$, along with $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ if the CDF of X is

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ \frac{x}{\ell}, & 0 \le x \le \ell, \\ 1, & x \ge \ell \end{cases}$$

(c) Suppose X and Y are independent and have densities

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

What is their joint distribution? (Hint: for parts (c) and (d), we can use independence in much the same way that we did in discrete probability)

(d) Calculate $\mathbb{E}[XY]$ for the X and Y in part (c).

Solution:

(a) Yes, it is a valid density function; it is non-negative and integrates to 1.

No, it is not a valid CDF; a CDF should go to 1 as x goes to infinity and be non-decreasing.

(b) We have

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = \begin{cases} \frac{1}{\ell}, & 0 \le x \le \ell \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \int_{x=0}^{\ell} x \cdot \frac{1}{\ell} \,\mathrm{d}x = \frac{\ell}{2}$$

$$\mathbb{E}[X^2] = \int_{x=0}^{\ell} x^2 \cdot \frac{1}{\ell} \,\mathrm{d}x = \frac{\ell^2}{3}$$

$$\mathrm{Var}(X) = \frac{\ell^2}{3} - \frac{\ell^2}{4} = \frac{\ell^2}{12}$$

This is known as the continuous uniform distribution over the interval $[0,\ell]$, sometimes denoted $\text{Uniform}[0,\ell]$.

(c) Note that due to independence,

$$\begin{aligned} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \mathbb{P}[X \in [x,x+\mathrm{d}x], Y \in [y,y+\mathrm{d}y]] \\ &= \mathbb{P}[X \in [x,x+\mathrm{d}x]]\mathbb{P}[Y \in [y,y+\mathrm{d}y]] \\ &\approx f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned}$$

so their joint distribution is $f(x,y) = 2x$ on the unit square $0 \le x \le 1$, $0 \le y \le 1$.

(d) We have

$$\mathbb{E}[XY] = \int_{x=0}^{1}\int_{y=0}^{1} xy \cdot 2x\,\mathrm{d}y\,\mathrm{d}x = \int_{x=0}^{1} x^2\,\mathrm{d}x = \frac{1}{3}.$$

Alternatively, since X and Y are independent, we can compute $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Note that

$$\mathbb{E}[X] = \int_{0}^{1} x \cdot 2x\,\mathrm{d}x = \frac{2}{3}x^3\bigg|_0^1 = \frac{2}{3},$$

and $\mathbb{E}[Y] = \frac{1}{2}$ since the density of Y is symmetric around $\frac{1}{2}$. Hence,

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{1}{3}.$$

2 Darts Again

Note 21

Edward and Khalil are playing darts on a circular dartboard.

Edward's throws are uniformly distributed over the entire dartboard, which has a radius of 10 inches. Khalil has good aim (but his throws may land outside of the dartboard); the distance of his throws from the center of the dartboard follows an exponential distribution with parameter $\frac{1}{2}$.

Say that Edward and Khalil both throw one dart at the dartboard. Let X be the distance of Edward's dart from the center, and Y be the distance of Khalil's dart from the center of the dartboard. What is $\mathbb{P}[X < Y]$, the probability that Edward's throw is closer to the center of the board than Khalil's? Leave your answer in terms of an unevaluated integral.

[*Hint:* X is not uniform over $[0,10]$. Solve for the distribution of X by first computing the CDF of X, $\mathbb{P}[X < x]$.]

Solution: We are given that $Y \sim \text{Exponential}(1/2)$. We now find the distribution of X by solving for the CDF of X, $\mathbb{P}[X < x]$. To get this, 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:

$$\mathbb{P}[X < x] = \frac{\pi x^2}{\pi 10^2} = \frac{x^2}{100}.$$

for $x \in (0,10)$. For $x < 0$, the CDF is 0, and for $x > 10$, the CDF is 1.

Differentiating gives us the PDF of X, which is given by $f_X(x) = \frac{x}{50}$ for $x \in (0,10)$, and 0 elsewhere. Now, we solve for $\mathbb{P}[X < Y]$ with total probability:

$$\begin{aligned}
\mathbb{P}[X < Y] &= \int_0^{10} \mathbb{P}[Y > X \mid X = x] f_X(x)\,\mathrm{d}x \\
&= \int_0^{10} \mathbb{P}[Y > x] f_X(x)\,\mathrm{d}x \\
&= \int_0^{10} e^{-0.5x}\frac{x}{50}\,\mathrm{d}x
\end{aligned}$$

$\mathbb{P}[Y > x] = e^{-0.5x}$ comes from the (complement of the) exponential CDF. Evaluating this integral gives us $\mathbb{P}[X < Y] \approx 0.0767$.

Explanation of Integral: The integral may seem a bit confusing, but let's break it down. This is an expression of the total probability rule, where we condition on X. Recall that in discrete, we could calculate

$\mathbb{P}[X < Y] = \sum_x \mathbb{P}[X = x]\mathbb{P}[Y > x]$. In continuous, it is pretty much analogous, just with an integral and $f_X(x)\,\mathrm{d}x$ instead of the summation of $\mathbb{P}[X = x]$.

Alternative Calculation of PDF: Another way we could've calculate the PDF of X is by noticing that the PDF corresponds to the likelihood of falling on a point with radius x. The portion of the circle that corresponds to the radius of x is equivalent to its circumference, which is linear with respect to the radius, thus the likelihood (PDF) should be linear with respect to the radius as well. Thus, we have that $f_X(x) = cx$ for $x \in (0,10)$, and 0 elsewhere. The PDF must integrate to 1, so $\int_0^{10} cx\,\mathrm{d}x = 50c = 1$, which means that $c = \frac{1}{50}$.

Alternative Setup of Integral: You may have noticed that we chose to condition on X in our setup for total probability. It happens to be that this is the easier way to setup the integral, because the bounds are simpler (since X is bounded by 0 to 10) and $\mathbb{P}[Y > X \mid X = x]$ has a simple, continuous expression. We could've instead conditioned on Y, but it is more difficult. The interested reader may choose to follow along with the alternate integral setup as follows.

$$\begin{aligned}
\mathbb{P}[X < Y] &= \int_0^{\infty} \mathbb{P}[X < Y \mid Y = y] f_Y(y)\,\mathrm{d}y \\
&= \int_0^{10} \frac{y^2}{100}\cdot 0.5e^{-y/2}\,\mathrm{d}y + \int_{10}^{\infty} 1\cdot 0.5e^{-y/2}\,\mathrm{d}y
\end{aligned}$$

This is because $\mathbb{P}[X < Y \mid Y = y]$ is the CDF of X: $F_X(y)$, which changes expression past $X = 10$. Thus, we must split the integral into two parts.

3 Lunch Meeting

Note 21

Alice and Bob agree to try to meet for lunch between 12 PM and 1 PM at their favorite sushi restaurant. Being extremely busy, they are unable to specify their arrival times exactly, and can say only that each of them will arrive (independently) at a time that is uniformly distributed within the hour. In order to avoid wasting precious time, if the other person is not there when they arrive they agree to wait exactly fifteen minutes before leaving.

(a) Provide a sketch of the joint distribution of the arrival times of Alice and Bob. For which region of the graph will Alice and Bob actually meet?

(b) Based on your sketch, what is the probability that they will actually meet for lunch?

Solution:

(a) Let the random variable A be the time that Alice arrives and the random variable B be the time when Bob arrives. Since A and B are both uniformly distributed, it is helpful to visualize the distribution graphically. Consider Figure 1, plotting the space of all outcomes (a, b):

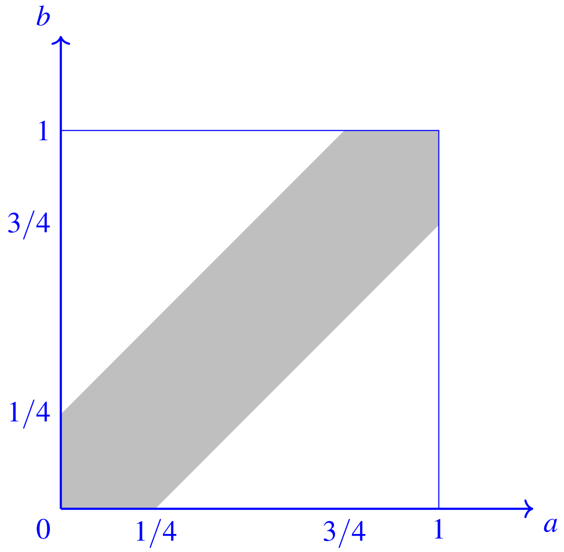


Figure 1: Visualization of joint probability density.

The arrival times are uniformly distributed over the box, and the shaded region is the set of values (a, b) for which Alice and Bob will actually meet for lunch.

(b) Since all points in this square are equally likely, the probably they meet is the ratio of the shaded area to the area of the square. If the area of the square is 1, then the area of the shaded region is

$$1 - 2 \times \left[\frac{1}{2} \times \left(\frac{3}{4}\right)^2\right] = \frac{7}{16},$$

since the area of the white triangle on the upper-left is $(1/2) \cdot (3/4)^2$, and the white triangle on the lower-right has the same area. Therefore, the probability that Alice and Bob actually meet is $7/16$.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

DIS 12B

1 Interesting Gaussians

Note 21

(a) If $X \sim N(0,\sigma_X^2)$ and $Y \sim N(0,\sigma_Y^2)$ are independent, then what is $\mathbb{E}\left[(X+Y)^k\right]$ for any *odd* $k \in \mathbb{N}$?

(b) Let $f_{\mu,\sigma}(x)$ be the density of a $N(\mu,\sigma^2)$ random variable, and let X be distributed according to $\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha)f_{\mu_2,\sigma_2}(x)$ for some $\alpha \in [0,1]$. Compute $\mathbb{E}[X]$ and $\operatorname{Var}(X)$. Is X normally distributed?

Solution:

(a) $\mathbb{E}\left[(X+Y)^k\right] = 0$.

Since X and Y are Gaussians, so must $Z = X+Y$ be. Specifically, $Z \sim N(0,\sigma_X^2+\sigma_Y^2)$. Thus, the PDF f_Z of Z is still symmetric about the origin; that is, it is an even function, i.e. $f_Z(x) = f_Z(-x)$ for any $a,b \in \mathbb{R}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[(X+Y)^k\right] = \mathbb{E}\left[Z^k\right] &= \int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{0} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{0}^{\infty} (-x)^k f_Z(-x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= -\int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= 0,
\end{aligned}$$

since k is odd.

Note that we could've just concluded that $\int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x = 0$ due to the fact that $x^k f_Z(x)$ is an odd function (since x^k is an odd function for odd k), and the integral from $(-a,a)$ for any odd function will evalute to 0.

Also note that adding two RVs is NOT equivalent to adding their PDFs. Instead, adding two RVs is equivalent to convolving their PDFs. As an example, for random variables $X+Y=Z$, it is true that $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx$.

(b) $\mathbb{E}[X] = \alpha\mu_1 + (1-\alpha)\mu_2$, $\operatorname{Var}(X) = \alpha\left(\sigma_1^2+\mu_1^2\right) + (1-\alpha)\left(\sigma_2^2+\mu_2^2\right) - (\mathbb{E}[X])^2$. No, X is not necessarily normally distributed.

$$\begin{aligned}
\mathbb{E}[X] := \mu &= \int_{-\infty}^{\infty} x\left(\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)\right) \mathrm{d}x \\
&= \alpha \int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x = \alpha\mu_1 + (1-\alpha)\mu_2 \\
\operatorname{Var}(X) := \sigma^2 &= \mathbb{E}\left[X^2\right] - \mu^2 = \alpha \int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x^2 f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x - \mu^2 \\
&= \alpha\left(\sigma_1^2 + \mu_1^2\right) + (1-\alpha)\left(\sigma_2^2 + \mu_2^2\right) - \mu^2.
\end{aligned}$$

We know that the density of $N(\mu,\sigma)$ has a unique maximum at $x = \mu$; however, if, e.g. $\alpha = 1/2, \mu_1 = -10, \mu_2 = 10, \sigma_1 = \sigma_2 = 1$, then $\alpha f_{\mu_1,\sigma_1} + (1-\alpha) f_{\mu_2,\sigma_2}$ has two maxima, and so cannot be the density of a Gaussian.

Explanation of integrals: $\int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}[X_1]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$, which is μ_1 by definition.

$\int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}\left[X_1^2\right]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$. $\mathbb{E}\left[X_1^2\right] = \operatorname{Var}(X_1) + \mathbb{E}[X_1]^2 = \sigma_1^2 + \mu_1^2$ by definition.

2 Binomial Concentration

Note 21

Here, we will prove that the binomial distribution is *concentrated* about its mean as the number of trials tends to ∞. Suppose we have i.i.d. trials, each with a probability of success $1/2$. Let S_n be the number of successes in the first n trials (n is a positive integer).

(a) Compute the mean and variance of S_n.

(b) How should we define Z_n in terms of S_n to ensure that Z_n has mean 0 and variance 1?

(c) What is the distribution of Z_n as $n \to \infty$?

(d) Use the bound $\mathbb{P}[Z > z] \le (\sqrt{2\pi}z)^{-1}\mathrm{e}^{-z^2/2}$ when Z is a standard normal in order to approximately bound $\mathbb{P}[S_n/n > 1/2 + \delta]$, where $\delta > 0$.

Solution:

(a) Since $S_n \sim \operatorname{Binomial}(n, \frac{1}{2})$, we have $\mathbb{E}[S_n] = \frac{n}{2}$ and $\operatorname{Var}(S_n) = \frac{n}{4}$.

(b) We can define

$$Z_n := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\operatorname{Var}(S_n)}} = \frac{S_n - n/2}{\sqrt{n}/2}.$$

In particular, we subtract the mean and divide by the standard deviation to normalize S_n.

To check, we have

$$\mathbb{E}[Z_n] = \frac{1}{\sqrt{n}/2}\mathbb{E}\left[S_n - \frac{n}{2}\right] = \frac{1}{\sqrt{n}/2}\left(\mathbb{E}[S_n] - \frac{n}{2}\right) = 0,$$
$$\mathrm{Var}(Z_n) = \frac{1}{n/4}\mathrm{Var}\left(S_n - \frac{n}{2}\right) = \frac{1}{n/4}\mathrm{Var}(S_n) = 1,$$

since $S_n \sim \mathrm{Binomial}(n, 1/2)$.

(c) The central limit theorem tells us that $Z_n \to \mathcal{N}(0,1)$.

(d) In order to apply the bound, we must apply it to Z_n.

$$\mathbb{P}\left[\frac{S_n}{n} > \frac{1}{2} + \delta\right] = \mathbb{P}\left[\frac{S_n - n/2}{n} > \delta\right] = \mathbb{P}\left[\frac{S_n - n/2}{\sqrt{n}/2} > 2\delta\sqrt{n}\right] \approx \mathbb{P}[Z_n > 2\delta\sqrt{n}]$$
$$\leq \frac{1}{2^{3/2}\delta\sqrt{\pi n}} e^{-2\delta^2 n}$$

3 Erasures, Bounds, and Probabilities

Note 21

Alice is sending 1000 bits to Bob. The probability that a bit gets erased is p, and the erasure of each bit is independent of the others.

Alice is using a scheme that can tolerate up to one-fifth of the bits being erased. That is, as long as Bob receives at least 801 of the 1000 bits correctly, he can decode Alice's message.

In other words, Bob becomes unable to decode Alice's message only if 200 or more bits are erased. We call this a "communication breakdown", and we want the probability of a communication breakdown to be at most 10^{-6}.

(a) Use Chebyshev's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6}.

(b) As the CLT would suggest, approximate the fraction of erasures by a Gaussian random variable (with suitable mean and variance). Use this to find an approximate bound for p such that the probability of a communications breakdown is at most 10^{-6}.

You may use that $\Phi^{-1}(1-10^{-6}) \approx 4.753$.

Solution:

(a) Let X be the random variable denoting the number of erasures. Chebyshev's inequality states the following:

$$\mathbb{P}[|X - \mu_X| \geq k] \leq \frac{\sigma_X^2}{k^2}.$$

This gives us the bound

$$\begin{aligned}\mathbb{P}[X \geq 200] &= \mathbb{P}[X - \mu_X \geq 200 - \mu_X] \\ &\leq \mathbb{P}[|X - \mu_X| \geq 200 - \mu_X] \\ &\leq \frac{\sigma_X^2}{(200 - \mu_X)^2}\end{aligned}$$

Since $X \sim \text{Binomial}(1000, p)$, we have $\mu_X = 1000p$ and $\sigma_X^2 = 1000p(1-p)$. Substituting these values in, we have

$$\mathbb{P}[X \geq 200] \leq \frac{1000p(1-p)}{(200 - 1000p)^2} = \frac{p(1-p)}{40(1-5p)^2}.$$

To meet our objective, we just have to ensure that

$$\mathbb{P}[X \geq 200] \leq \frac{p(1-p)}{40(1-5p)^2} \leq 10^{-6},$$

which yields an upper bound of about 3.998×10^{-5} for p.

(b) Let Y be equal to the fraction of erasures, i.e. $\frac{X}{1000}$. Using properties of expectation and variance, we can see that

$$\begin{aligned}\mathbb{E}[Y] &= p \\ \text{Var}(Y) &= \text{Var}(X) \cdot \frac{1}{1000^2} = \frac{p(1-p)}{1000}\end{aligned}$$

Therefore, by Central Limit Theorem, we can say that Y is roughly a normal distribution with that mean and variance. Since we are interested in the event that $Y \geq 0.2$, let's figure out how many standard deviations above the mean 0.2 is:

$$\frac{0.2 - p}{\sqrt{\frac{p(1-p)}{1000}}} = \frac{(0.2 - p)\sqrt{1000}}{\sqrt{p(1-p)}}.$$

Therefore, the probability that we get a failure should be approximately (by CLT),

$$1 - \Phi\left(\frac{(0.2 - p)\sqrt{1000}}{\sqrt{p(1-p)}}\right)$$

where Φ is the CDF of a standard normal variable. Setting this to be at most 10^{-6} gives us

$$\Phi\left(\frac{(0.2 - p)\sqrt{1000}}{\sqrt{p(1-p)}}\right) \geq 1 - 10^{-6}$$

And, since $\Phi^{-1}(1 - 10^{-6}) \approx 4.753$, we solve the inequality

$$\frac{(0.2 - p)\sqrt{1000}}{\sqrt{p(1-p)}} \geq 4.753$$

This yields that we need $p \leq 0.1468$.

Note that this gives quite a different value from the previous parts. This is because the Central Limit Theorem gives a much tighter approximation for tail events than Markov's and Chebyshev's. However, we can only apply the Central Limit Theorem because n is large.

Therefore, we do not need p to be so low to achieve a communication breakdown probability of 10^{-6}. The other bounds required us to need a probability of on the order of 10^{-5}, but here we realize that we only need it to be less than 0.1468. (The true bound is .1459.) Quite drastic!

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair

Discussion 13A

1 Markov Chain Basics

Note 22

A Markov chain is a sequence of random variables X_n, $n = 0, 1, 2, \dots$. Here is one interpretation of a Markov chain: X_n is the state of a particle at time n. At each time step, the particle can jump to another state. Formally, a Markov chain satisfies the Markov property:

$$\mathbb{P}[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = \mathbb{P}[X_{n+1} = j \mid X_n = i], \tag{1}$$

for all n, and for all sequences of states $i_0, \dots, i_{n-1}, i, j$. In other words, the Markov chain does not have any memory; the transition probability only depends on the current state, and not the history of states that have been visited in the past.

(a) In lecture, we learned that we can specify Markov chains by providing three ingredients: $\mathscr{X}$, P, and π_0. What do these represent, and what properties must they satisfy?

(b) If we specify $\mathscr{X}$, P, and π_0, we are implicitly defining a sequence of random variables X_n, $n = 0, 1, 2, \dots$, that satisfies (1). Explain why this is true.

(c) Calculate $\mathbb{P}[X_1 = j]$ in terms of π_0 and P. Then, express your answer in matrix notation. What is the formula for $\mathbb{P}[X_n = j]$ in matrix form?

Solution:

(a) $\mathscr{X}$ is the set of states, which is the range of possible values for X_n. In this course, we only consider finite $\mathscr{X}$.

P contains the transition probabilities. $P(i, j)$ is the probability of transitioning from state i to state j. It must satisfy $\sum_{j \in \mathscr{X}} P(i, j) = 1 \ \forall i \in \mathscr{X}$, which says that the probability that *some* transition occurs must be 1. Also, the entries must be non-negative: $P(i, j) \geq 0 \ \forall i, j \in \mathscr{X}$. A matrix satisfying these two properties is called a stochastic matrix.

Note that we allow states to transition to themselves, i.e. it is possible for $P(i, i) > 0$.

π_0 is the initial distribution, that is, $\pi_0(i) = \mathbb{P}[X_0 = i]$. Similarly, we let π_n be the distribution of X_n. Since π_0 is a probability distribution, its entries must be non-negative and $\sum_{i \in \mathscr{X}} \pi_0(i) = 1$.

(b) The sequence of random variables X_n, $n = 0, 1, 2, \dots$, is defined in the following way:

- X_0 has distribution π_0, i.e. $\mathbb{P}[X_0 = i] = \pi_0(i)$.
- X_{n+1} has distribution given by

$$\mathbb{P}[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = \mathbb{P}[X_{n+1} = j \mid X_n = i] = P(i, j),$$

for all $n = 0, 1, 2, \dots$.

It is important to realize the connection between the Markov property (1) and the transition matrix P. P contains information about the transition probabilities in one step. If the Markov property did not hold, then P would not be enough to specify the distribution of X_{n+1}. Conversely, if we only specify P, then we are implicitly assuming that the transition probabilities do not depend on anything other than the current state. Note that this convention is different from what EE16A uses, if you have taken that class/are taking it right now.

(c) By the Law of Total Probability,

$$\mathbb{P}[X_1 = j] = \sum_{i\in\mathscr{X}} \mathbb{P}[X_1 = j, X_0 = i] = \sum_{i\in\mathscr{X}} \mathbb{P}[X_0 = i]\mathbb{P}[X_1 = j \mid X_0 = i] = \sum_{i\in\mathscr{X}} \pi_0(i)P(i,j).$$

If we write $\pi_1(j) = \mathbb{P}[X_1 = j]$ and π_0 as row vectors, then in matrix notation we have

$$\pi_1 = \pi_0 P.$$

The effect of a transition is right-multiplication by P. After n time steps, we have

$$\pi_n = \pi_0 P^n.$$

At this point, it should be mentioned that many calculations involving Markov chains are very naturally expressed with the language of matrices. Consequently, Markov chains are very well-suited for computers, which is one of the reasons why Markov chain models are so popular in practice.

2 Can it be a Markov Chain?

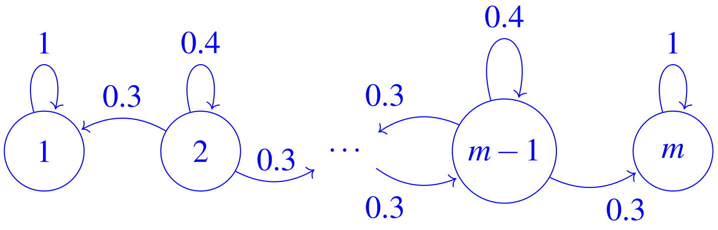
Note 22

(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 the 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 memoryless property is violated.

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

Note 22

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 exactly one 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.

Solution:

(a) Let state i represent the situation that Allen has i umbrellas at his current location, for $i = 0$, 1, or 2.

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,

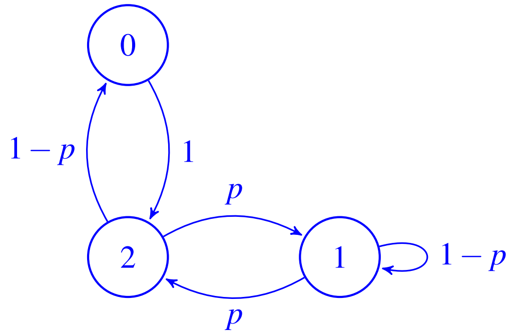
$$\mathbb{P}[X_{n+1} = 2 \mid X_n = 0] = 1.$$

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:

$$\begin{bmatrix} \pi(0) & \pi(1) & \pi(2) \end{bmatrix} = \frac{1}{3-p} \begin{bmatrix} 1-p & 1 & 1 \end{bmatrix}$$

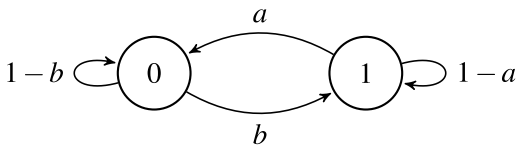
CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair **Discussion 13B**

1 Markov Chain Terminology

Note 22

In this question, we will walk you through terms related to Markov chains.

1. (Irreducibility) A Markov chain is irreducible if, starting from any state i, the chain can transition to any other state j, possibly in multiple steps.
2. (Periodicity) $d(i) := \gcd\{n > 0 \mid P^n(i,i) = \mathbb{P}[X_n = i \mid X_0 = i] > 0\}, i \in \mathscr{X}$. If $d(i) = 1 \; \forall i \in \mathscr{X}$, then the Markov chain is aperiodic; otherwise it is periodic.
3. (Matrix Representation) Define the transition probability matrix P by filling entry (i, j) with probability $P(i, j)$.
4. (Invariance) A distribution π is invariant for the transition probability matrix P if it satisfies the following balance equations: $\pi = \pi P$.



(a) For what values of a and b is the above Markov chain irreducible? Reducible?

(b) For $a = 1$, $b = 1$, prove that the above Markov chain is periodic.

(c) For $0 < a < 1$, $0 < b < 1$, prove that the above Markov chain is aperiodic.

(d) Construct a transition probability matrix using the above Markov chain.

(e) Write down the balance equations for this Markov chain and solve them. Assume that the Markov chain is irreducible.

Solution:

(a) The Markov chain is irreducible if both a and b are non-zero. It is reducible if at least one of a and b is 0.

(b) We compute $d(0)$ to find that:

$$d(0) = \gcd\{2, 4, 6, \ldots\} = 2.$$

This is because if we start at a state X then we can get back to it after taking an even number of steps only (2, 4, 6, 8, etc.), not by taking an odd number of steps (1, 3, 5, 7, etc.). Thus, the chain is periodic with period 2.

(c) We compute $d(0)$ to find that:

$$d(0) = \gcd\{1,2,3,...\} = 1.$$

Thus, the chain is aperiodic. Notice that the self-loops allow us to stay at the same node, thereby letting us get to any other node in an odd *or* even number of steps.

(d)

$$\begin{bmatrix} 1-b & b \\ a & 1-a \end{bmatrix}$$

(e)

$$\pi(0) = (1-b)\pi(0) + a\pi(1),$$
$$\pi(1) = b\pi(0) + (1-a)\pi(1).$$

One of the equations is redundant. We throw out the second equation and replace it with $\pi(0) + \pi(1) = 1$. This gives the solution

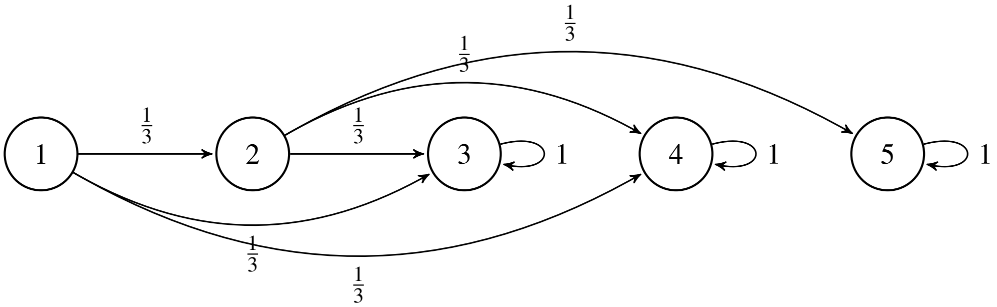
$$\pi = \frac{1}{a+b}\begin{bmatrix} a & b \end{bmatrix}.$$

2 Skipping Stones

Note 22

We consider a simple Markov chain model for skipping stones on a river, but with a twist: instead of trying to make the stone travel as far as possible, you want the stone to hit a target. Let the set of states be $\mathscr{X} = \{1,2,3,4,5\}$. State 3 represents the target, while states 4 and 5 indicate that you have overshot your target. Assume that from states 1 and 2, the stone is equally likely to skip forward one, two, or three steps forward. If the stone starts from state 1, compute the probability of reaching our target before overshooting, i.e. the probability of $\{3\}$ before $\{4,5\}$.

Solution: Here is the Markov Chain we are working with:



Let $\alpha(i)$ denote the probability of reaching the target before overshooting, starting at state i. Then:

$$\begin{aligned}
\alpha(5) &= 0\\
\alpha(4) &= 0\\
\alpha(3) &= 1\\
\alpha(2) &= \frac{1}{3}\alpha(3) + \frac{1}{3}\alpha(4) + \frac{1}{3}\alpha(5) = \frac{1}{3}\\
\alpha(1) &= \frac{1}{3}\alpha(2) + \frac{1}{3}\alpha(3) + \frac{1}{3}\alpha(4) = \frac{1}{9} + \frac{1}{3}
\end{aligned}$$

Therefore, $\alpha(1) = 1/9 + 1/3 = 4/9$.

3 Consecutive Flips

Note 22

Suppose you are flipping a fair coin (one Head and one Tail) until you get the same side 3 times (Heads, Heads, Heads) or (Tails, Tails, Tails) in a row.

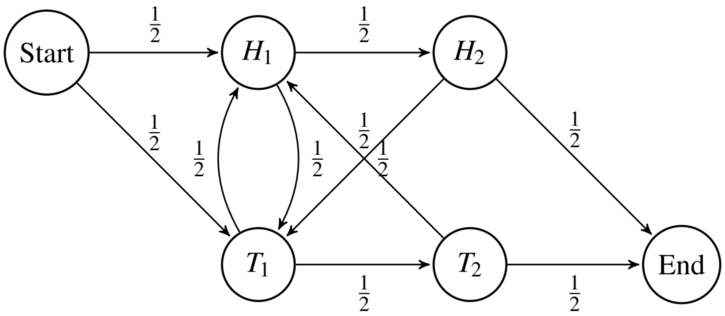
(a) Construct an Markov chain that describes the situation with a start state and end state.

(b) Given that you have flipped a (Tails, Heads) so far, what is the expected number of flips to see the same side three times?

(c) What is the expected number of flips to see the same side three times, beginning at the start state?

Solution:

(a) The appropriate Markov chain has 6 states: Start, H_1, H_2, T_1, T_2, and End.
For starting node, there is an outgoing edge to H_1 and T_1, each with equal probability of $1/2$.
For H_1, there is an outgoing edge to H_2 and T_1, each with equal probability of $1/2$.
For H_2, there is an outgoing edge to End and T_1, each with equal probability of $1/2$.
For T_1, there is an outgoing edge to H_1 and T_2, each with equal probability of $1/2$.
For T_2, there is an outgoing edge to H_1 and End, each with equal probability of $1/2$.



(b) If you got a Tails and then a Heads, you are currently in the H_1 state. Thus, we must calculate the expected number of flips to end from H_1. Thus we will do this with a system of equations. Since we are not trying to solve for the starting state, we have 5 unknowns that depend on 5 linearly independent equations. Let $\beta(i)$ be the expected number of flips to reach the end state starting from state i. Then we have:

$$\begin{aligned}
\beta(H_1) &= 1 + 0.5\beta(T_1) + 0.5\beta(H_2) \\
\beta(H_2) &= 1 + 0.5\beta(\text{End}) + 0.5\beta(T_1) \\
\beta(T_1) &= 1 + 0.5\beta(T_2) + 0.5\beta(H_1) \\
\beta(T_2) &= 1 + 0.5\beta(\text{End}) + 0.5\beta(H_1) \\
\beta(\text{End}) &= 0
\end{aligned}$$

If we solve this system of equations, we get $\beta(H_1) = 6, \beta(H_2) = 4, \beta(T_1) = 6, \beta(T_2) = 4$.

(c) $\beta(S) = 1 + 0.5\beta(H_1) + 0.5\beta(T_1) = 1 + 0.5 \cdot 6 + 0.5 \cdot 6 = 7$.

CS 70 Discrete Mathematics and Probability Theory
Spring 2024 Seshia, Sinclair **Discussion 14**

1 Short Tree Proofs

Note 5

Let $G = (V, E)$ be an undirected graph with $|V| \geq 1$.

(a) Prove that every connected component in an acyclic graph is a tree.

(b) Suppose G has k connected components. Prove that if G is acyclic, then $|E| = |V| - k$.

(c) Prove that a graph with $|V|$ edges contains a cycle.

Solution:

(a) Every connected component is connected, and acyclic because the graph is acyclic; by definition, this is a tree.

(b) Because each connected component is a tree, each connected component has $|V_i| - 1$ edges. The total number of edges is thus $\sum_i(|V_i| - 1) = |V| - k$.

(c) An acyclic graph has $|V| - k$ edges which cannot equal $|V|$, thus if a graph has $|V|$ edges it has a cycle.

2 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.

3 Counting Subsets

Note 11

Consider the set S of all (possibly infinite) subsets of $\mathbb{N}$.

(a) Show that there is a bijection between S and $T = \{f : \mathbb{N} \to \{0,1\}\}$ (the set of all functions that map each natural number to 0 or 1).

(b) Prove or disprove: S is countable.

(c) Say that a function $f : \mathbb{N} \to \{0,1\}$ has *finite support* if it is non-zero on only a finite set of inputs. Let F denote the set of functions $f : \mathbb{N} \to \{0,1\}$ with finite support.

Prove that F is countably infinite.

Solution:

(a) Let $X \subseteq \mathbb{N}$. Define $f(x) = 1$ if $x \in X$, and 0 otherwise.

This is onto, since for a function f, the set that maps to it is $\{x \in \mathbb{N} \mid f(x) = 1\}$.

This is one-to-one. By showing the contrapositive, if two different set X and X' differ on some value x. Let $x \in X$ and $x \notin X'$. The function f and f' which they map to respectively will differ at $f(x) = 1$ and $f'(x) = 0$. Thus $f \neq f'$.

(b) Uncountable. Note that such f can be viewed as a binary encoding of a real number between 0 and 1, which exhibits a surjection from V to $[0,1]$.

(c) We give a bijection between F and $\mathbb{N}$. We encode an $f \in S$ as a binary number y_f, with the ith position (with $i = 0$ being the least significant digit) set to 1 if $f(i) = 1$. Note that this encoding always has finite length, excluding leading zeros, since the maximum i for which $f(i) = 1$ is finite. Thus the encoding always results in a natural number encoded in binary. This conversion is one-to-one, since each f and f' in F differ on at least one input and therefore y_f and $y_{f'}$ differ in at least one position. The conversion is onto, since every binary number represents a function with finite support. Since the natural numbers are countably infinite, and we have a bijection between F to $\mathbb{N}$, F is countably infinite.

An alternative bijection is between F and the subset of $\mathbb{N}$ that contains only numbers that are the product of distinct primes. Let $\{p_0 = 2, p_1 = 3, p_2 = 5, \ldots\}$ be the set of all primes where

p_i is the $(i+1)$th prime. As shown in a previous homework, this set is infinite. Now consider a function $f(x)$ that is 1 exactly on inputs $x_1, x_2, \ldots, x_k$. Encode $f(x)$ as the natural number $p_{x_1} \times p_{x_2} \times \cdots \times p_{x_k}$. In other words, the function f is encoded as the natural number $2^{f(0)} \times 3^{f(1)} \times 5^{f(2)} \times 7^{f(3)} \times 11^{f(4)} \times \cdots$. This encoding is one-to-one, since the prime factorization of a number is unique. The encoding is onto, since every natural number that is composed of distinct primes corresponds to a function in F. Thus this is a bijection, and F is countable.

4 Strings

Note 10

How many different strings of length 5 only contain A, B, C? And how many such strings contain at least one of each characters?

Solution: The number of different strings of length 5 is 3^5 since each position have 3 different choices.

Let E_A be the set of strings that the character A is not used in the string. We define E_B, E_C similarly. Then the total number of "bad" strings is $|E_A \cup E_B \cup E_C|$.

By the Principle of Inclusion and Exclusion,

$$|E_A \cup E_B \cup E_C| = |E_A| + |E_B| + |E_C| - |E_A \cap E_B| - |E_A \cap E_C| - |E_B \cap E_C| + |E_A \cap E_B \cap E_C| = 3 \cdot 2^5 - 3 \cdot 1 = 93$$

where $|E_A \cap E_B| = |E_B \cap E_C| = |E_C \cap E_A| = 1$, and $|E_A \cap E_B \cap E_C| = 0$. Thus, the total number of valid string is $3^5 - 93 = 150$

5 Mario's Coins

Note 14

Mario owns three identical-looking coins. One coin shows heads with probability $1/4$, another shows heads with probability $1/2$, and the last shows heads with probability $3/4$.

(a) Mario randomly picks a coin and flips it. He then picks one of the other two coins and flips it. Let X_1 and X_2 be the events of the 1st and 2nd flips showing heads, respectively. Are X_1 and X_2 independent? Please prove your answer.

(b) Mario randomly picks a single coin and flips it twice. Let Y_1 and Y_2 be the events of the 1st and 2nd flips showing heads, respectively. Are Y_1 and Y_2 independent? Please prove your answer.

(c) Mario arranges his three coins in a row. He flips the coin on the left, which shows heads. He then flips the coin in the middle, which shows heads. Finally, he flips the coin on the right. What is the probability that it also shows heads?

Solution:

(a) X_1 and X_2 are not independent. Intuitively, the fact that X_1 happened gives some information about the first coin that was chosen; this provides some information about the second coin that was chosen (since the first and second coins can't be the same coin), which directly affects whether X_2 happens or not.

To make this formal, we compute

$$\mathbb{P}[X_1] = \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{4}\right) = \frac{1}{2}$$

By symmetry, $\mathbb{P}[X_2] = \mathbb{P}[X_1]$, so

$$\mathbb{P}[X_1]\,\mathbb{P}[X_2] = \frac{1}{4}.$$

But if we consider the probability that both X_1 and X_2 happen, we have

$$\begin{aligned}\mathbb{P}[X_1 \cap X_2] &= \frac{1}{6}\Big[\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) + \\ &\qquad \left(\frac{1}{2}\right)\left(\frac{3}{4}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{2}\right)\Big] \\ &= \frac{22}{96} = \frac{11}{48}\end{aligned}$$

which is not equal to $1/4$, violating the definition of independence.

(b) Y_1 and Y_2 are not independent. Intuitively, the fact that Y_1 happens gives some information about the coin that was picked, which directly influences whether Y_2 happens or not.

To make this formal, we compute

$$\mathbb{P}[Y_1] = \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{4}\right) = \frac{1}{2}$$

By symmetry, $\mathbb{P}[Y_2] = \mathbb{P}[Y_1]$, so

$$\mathbb{P}[Y_1]\,\mathbb{P}[Y_2] = \frac{1}{4}$$

But if we consider the probability that both Y_1 and Y_2 happen, we have

$$\mathbb{P}[Y_1 \cap Y_2] = \left(\frac{1}{3}\right)\left(\frac{1}{4}\right)^2 + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)^2 + \left(\frac{1}{3}\right)\left(\frac{3}{4}\right)^2 = \frac{14}{48} = \frac{7}{24}$$

which is not equal to $1/4$, violating the definition of independence.

(c) Let A be the coin with bias $1/4$, B be the fair coin, and C be the coin with bias $3/4$. There are six orderings, each with probability $1/6$: ABC, ACB, BAC, BCA, CAB, and CBA. Thus

$$\begin{aligned}&\mathbb{P}[\text{Third coin shows heads} \mid \text{First two coins show heads}] \\ &= \frac{\mathbb{P}[\text{All three coins show heads}]}{\mathbb{P}[\text{First two coins show heads}]} \\ &= \frac{(\frac{1}{4})(\frac{1}{2})(\frac{3}{4})}{11/48} \\ &= \frac{3/32}{11/48} = \frac{9}{22}.\end{aligned}$$

Note that the denominator was the probability calculated in part a, so we just plugged it in as $\frac{11}{48}$.

6 Sum of Poisson Variables

Note 19

Assume that you were given two independent Poisson random variables X_1, X_2. Assume that the first has mean λ_1 and the second has mean λ_2. Prove that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Hint: Recall the binomial theorem.

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

Solution:

To show that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$, we have show that

$$\mathbb{P}[(X_1 + X_2) = i] = \frac{(\lambda_1 + \lambda_2)^i}{i!} e^{-(\lambda_1 + \lambda_2)}.$$

We proceed as follows:

$$\begin{aligned}
\mathbb{P}[(X_1 + X_2) = i] &= \sum_{k=0}^{i} \mathbb{P}[X_1 = k, X_2 = (i-k)] = \sum_{k=0}^{i} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \cdot \frac{\lambda_2^{i-k}}{(i-k)!} e^{-\lambda_2} \\
&= e^{-\lambda_1} e^{-\lambda_2} \sum_{k=0}^{i} \frac{1}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} = \frac{e^{-\lambda_1} e^{-\lambda_2}}{i!} \sum_{k=0}^{i} \frac{i!}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} \\
&= \frac{e^{-(\lambda_1 + \lambda_2)}}{i!} \sum_{k=0}^{i} \binom{i}{k} \lambda_1^k \lambda_2^{i-k} = \frac{e^{-(\lambda_1 + \lambda_2)}}{i!} (\lambda_1 + \lambda_2)^i
\end{aligned}$$

In the last line, we use the binomial expansion.

7 Balls in Bins

Note 15

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 a ball lands in a nonempty 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).

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

(b) Let I_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

$$\mathbb{E}[I_1 + \ldots + I_n] = \mathbb{E}[I_1] + \ldots + \mathbb{E}[I_n] = n\left(\frac{n-1}{n}\right)^k$$

(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.

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

8 Inequality Practice

Note 17

(a) X is a random variable such that $X \geq -5$ 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 $Y \leq 10$ 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

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

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

Finally, we note that we can upper bound $\mathbb{P}[|Z - 350| > 50]$ with $\mathbb{P}[|Z - 350| \geq 50]$.

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

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

9 Exponential Distributions: Lightbulbs

Note 21

A brand new lightbulb has just been installed in our classroom, and you know the life span of a lightbulb is exponentially distributed with a mean of 50 days.

(a) Suppose an electrician is scheduled to check on the lightbulb in 30 days and replace it if it is broken. What is the probability that the electrician will find the bulb broken?

(b) Suppose the electrician finds the bulb broken and replaces it with a new one. What is the probability that the new bulb will last at least 30 days?

(c) Suppose the electrician finds the bulb in working condition and leaves. What is the probability that the bulb will last at least another 30 days?

Solution:

(a) Let $X \sim \text{Exponential}(1/50)$ be the time until the bulb is broken. For an exponential random variable with parameter λ, the density function is $f_X(x) = \lambda\, \mathrm{e}^{-\lambda x}$ for $x > 0$. So in this case

$\lambda = 1/50$. Thus we can integrate the density to find the probability that the lightbulb broke in the first 30 days:

$$\mathbb{P}[X < 30] = \int_0^{30} \left(\frac{1}{50} \cdot \mathrm{e}^{-x/50}\right) \mathrm{d}x = 1 - \mathrm{e}^{-30/50} = 1 - \mathrm{e}^{-3/5} \approx 0.451.$$

(b) The new bulb's waiting time Y is i.i.d. with the old bulb's. So the answer is

$$\mathbb{P}[Y > 30] = 1 - \mathbb{P}[Y < 30] = 1 - (1 - \mathrm{e}^{-3/5}) = \mathrm{e}^{-3/5} \approx 0.549.$$

(c) The bulb is memoryless, so the probability it will last 60 days given that it has lasted 30 days, is just the probability it will last 30 days:

$$\mathbb{P}[X > 60 \mid X > 30] = \mathbb{P}[X - 30 > 30 \mid X > 30] = \mathbb{P}[X > 30] = \mathrm{e}^{-3/5} \approx 0.549.$$

10 Continuous Probability Continued

Note 17

For the following questions, please briefly justify your answers or show your work.

(a) Assume $\text{Bob}_1, \text{Bob}_2, \dots, \text{Bob}_k$ each hold a fair coin whose two sides show numbers instead of heads and tails, with the numbers on Bob_i's coin being i and $-i$. Each Bob tosses their coin n times and sums up the numbers he sees; let's call this number X_i. For large n, what is the distribution of $(X_1 + \cdots + X_k)/\sqrt{n}$ approximately equal to?

(b) If $X_1, X_2, \dots$ is a sequence of i.i.d. random variables of mean μ and variance σ^2, what is $\lim_{n\to\infty} \mathbb{P}\left[\sum_{k=1}^n \frac{X_k - \mu}{\sigma n^\alpha} \in [-1,1]\right]$ for $\alpha \in [0,1]$ (your answer may depend on α and Φ, the CDF of a $N(0,1)$ variable)?

Solution:

(a) $\boxed{N\left(0, \sum_{i=1}^k i^2\right)}$.

$(X_1 + \cdots + X_k)/\sqrt{n} = \frac{X_1}{\sqrt{n}} + \cdots + \frac{X_k}{\sqrt{n}}$, and since each $\frac{X_i}{\sqrt{n}}$ converges to $N\left(0, i^2\right)$ by the central limit theorem, their sum must converge to $N\left(0, \sum_{i=1}^k i^2\right)$. Alternatively, if we let X_j^i be the j^{th} coin toss of Bob_i, then $(X_1 + \cdots + X_k)/\sqrt{n} = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j^1 + \cdots + X_j^k)$. But the $Y_j = X_j^1 + \dots X_j^k$ themselves are i.i.d. variables of mean 0 and variance $\sum_{i=1}^k i^2$, and so the central limit theorem again implies a limiting distribution of $N\left(0, \sum_{i=1}^k i^2\right)$ (this constitutes an alternative proof of the fact that the sum of Gaussians is also a Gaussian, which we showed in class).

(b) $$\boxed{\lim_{n\to\infty} \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \begin{cases} 1, & \text{if } \alpha > \frac{1}{2}, \\ \Phi(1) - \Phi(-1), & \text{if } \alpha = \frac{1}{2}, \\ 0, & \text{if } \alpha < \frac{1}{2} \end{cases}}.$$

For $\alpha > \frac{1}{2}$, the reasoning is exactly as in the law of large numbers: By Chebyshev's inequality, we have $1 - \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \notin [-1,1]\right] \leq \frac{1}{n^{2\alpha-1}} \xrightarrow{n\to\infty} 0$. The $\alpha = \frac{1}{2}$ case is a direct consequence of the central limit theorem, while the $\alpha < \frac{1}{2}$ case follows indirectly from it: $\mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma\sqrt{n}} \in \left[-\frac{1}{n^{\frac{1}{2}-\alpha}}, \frac{1}{n^{\frac{1}{2}-\alpha}}\right]\right]$ $\approx \mathbb{P}\left[N(0,1) \in \left[-\frac{1}{n^{\frac{1}{2}-\alpha}}, \frac{1}{n^{\frac{1}{2}-\alpha}}\right]\right] \xrightarrow{n\to\infty} 0$.

11 Three Tails

Note 22

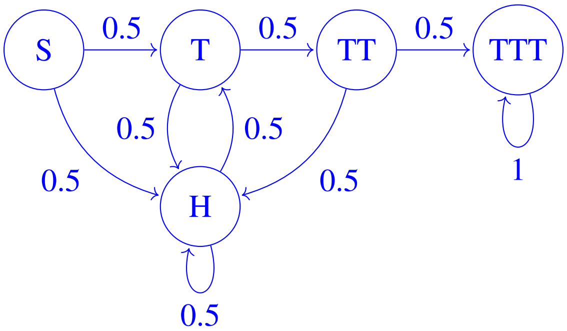
You flip a fair coin until you see three tails in a row. What is the average number of heads that you'll see until getting TTT?

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:

$$
\begin{aligned}
\beta(S) &= 0.5\beta(T) + 0.5\beta(H) && (1)\\
\beta(H) &= 1 + 0.5\beta(H) + 0.5\beta(T) && (2)\\
\beta(T) &= 0.5\beta(TT) + 0.5\beta(H) && (3)\\
\beta(TT) &= 0.5\beta(H) + 0.5\beta(TTT) && (4)\\
\beta(TTT) &= 0 && (5)
\end{aligned}
$$

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.