CS201A/201: Math for CS I/Discrete Mathematics midsem

`Max marks: 110`
`Time: 180 mins.` 22-Nov-2014

1. *Answer all 6 questions worth 110 marks. The paper has 4 pages.*
2. **Please start an answer to a question on a fresh page. And keep answers of parts of a question together.**
3. *Please show all calculations/justifications. No credit for only the final answer.*
4. *You can consult* **only your own handwritten notes**. *Nothing else is allowed.*

1. (a) i. Two cards are dealt from a properly shuffled pack. What is the probability that the second card is a heart ($\heartsuit$)? Clearly state any assumptions that you make in your calculation.

Solution:
Since we have no other information and there are 4 suits that are equally likely the probability is $\frac{1}{4}$. See 1a)iii) for an explanation.

ii. Supposing instead of two cards as in part i), five cards are dealt. Then what is the probablity that the 5^{th} card is a heart ($\heartsuit$)? Again clearly state any assumptions that you make in your calculation.

Solution:
Again as in 1a)i) since there is no extra information there are 4 equally likely suits and the probability is $\frac{1}{4}$. This seems counter-intuitive but see 1a)iii) why this is right.

iii. Can you intuitively explain the relationship that you observe between the answers to part 1a)i) and part 1a)ii).

Solution:
The fact that the probability does not change whether we deal 2 cards or 5 cards seems counter-intuitive. But given that there is no other information and there are 4 equally likely suits the probability is always $\frac{1}{4}$ irrespective of how many cards are dealt. We can calculate this the hard way and confirm that we get the same answer.
Let H_n be the event that the n^{th} card dealt is a heart and H_n^c be the event that the n^{th} card dealt is not a heart. Then for 1a)i):

$$P(H_2) = P(H_1 \cap H_2) + P(H_1^c \cap H_2) = \frac{13}{52} \times \frac{12}{51} + \frac{39}{52} \times \frac{13}{51} = \frac{1}{4}$$

We are assuming that H_1 and H_2 are independent which is reasonable. This also makes H_1^c and H_2 independent.
Note that this way becomes much harder for 1a)ii) since now there 2^4 possibilities for the first 4 cards being a heart or not. This will shoot up to 2^{m-1} when we are looking at m card deals.
One way to shorten this calculation is to assume that all 5-card hands are equally likely and calculate the number of hands that have a heart as the 5^{th} card. The number of 5-card hands is: $52 \times 51 \times 50 \times 49 \times 48$. If we fix the 5^{th} card the number of 5-card hands is: $51 \times 50 \times 49 \times 48$. There are 13 cards in the heart ($\heartsuit$) suit, so we have to multiply the previous number by 13 to get the number of 5-card hands with the 5^{th} card as a heart. This gives a probability of:

$$\frac{13 \times 51 \times 50 \times 49 \times 48}{52 \times 51 \times 50 \times 49 \times 48} = \frac{13}{52} = \frac{1}{4}$$

Note how this argument works for any m-card deal.

(b) A document that you need is equally likely to be in one of 3 files F_1, F_2, F_3. Let α_1, α_2, α_3; $\alpha_i < 1,\ i = 1..3$ be respectively the probabilities that you will find the document when you hurriedly go through files F_1, F_2, F_3 when the document is in fact in file F_i. Suppose you look in F_1 and do not find the document. What is the probablity it is in F_1? First, carefully define the events and then calculate the required probability.

Solution:
Let $E_1,\ E_2,\ E_3$ be the events that the document is in files F_1, F_2, F_3 respectively - note that these are mutually exclusive. Let E be the event that a hurried search in folder F_1 does not find the document. We wish to calculate the probability $P(E_1|E)$. We can use Bayes rule to write: $P(E_1|E) = \frac{P(E|E_1)P(E_1)}{P(E)}$.
By the rule of total probability we can write $P(E)$ as:

$$\begin{aligned} P(E) &= P(E \cap E_1) + P(E \cap E_2) + P(E \cap E_3) \\ &= P(E|E_1)P(E_1) + P(E|E_2)P(E_2) + P(E|E_3)P(E_3) \\ &= (1 - \alpha_1) \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3} \\ &= \frac{(3 - \alpha_1)}{3} \end{aligned}$$

So,

$$P(E_1|E) = \frac{(1 - \alpha_1) \times \frac{1}{3}}{(3 - \alpha_1) \times \frac{1}{3}} = \frac{(1 - \alpha_1)}{(3 - \alpha_1)}.$$

[(3,3,4),5=15]

2. (a) Can $f(x) = \frac{tan^{-1}(x)}{\pi} + \frac{1}{2}$, $-\infty < x < \infty$ be a cdf? Please justify. No marks for writing just yes/no.

Solution:
It is a cdf. We check that f satisifies the 3 conditions for a cdf.
1) $lim_{x\to-\infty}\frac{tan^{-1}(x)}{\pi} + \frac{1}{2} = \frac{-\pi/2}{\pi} + \frac{1}{2} = 0$. Also, $lim_{x\to\infty}\frac{tan^{-1}(x)}{\pi} + \frac{1}{2} = \frac{\pi/2}{\pi} + \frac{1}{2} = 1$.
2) & 3) Differentiating $f(x)$ we get $f'(x) = \frac{1}{\pi(1+x^2)}$. We see that the slope is positive for $-\infty < x < \infty$. So $f(x)$ is non-decreasing and continuous (so right continuous) in the given interval.

(b) A fair 6-sided die is rolled till each of the 6 outcomes has occurred at least once. Let RV (random variable) X be the number of rolls required. Find the probability mass function (pmf) of X.

Solution:
It is important to carefully define events else the calculation becomes messy.
E - desired event: all 6 faces appear at least once in r throws of the die.
E_i, $i = 1..6$: face i appears at least once in r throws of the die.
E_i^c: face i does not appear at all in r throws of the die.
X: RV denoting number of throws within which all 6 faces appear at least once.
$E = \cap_{i=1}^6 E_i = (\cup_{i=1}^6 E_i^c)^c = \Omega \setminus \cup_{i=1}^6 E_i^c$.
Using the inclusion-exclusion principle for probabilities (similar to that for sets, that is events).

$$P((\cup_{i=1}^6 E_i^c)^c)$$

$$= 1 - \sum_{k=1}^{6}(-1)^{k+1} \sum_{\substack{I\subset 1..6 \\ |I|=k}} P(E_I) \quad \text{where } E_I = \cap_{i\in I} E_i^c.$$

In our case the inner sum is $\binom{6}{k}(1-\frac{k}{6})^r$.

$$= 1 - \sum_{k=1}^{6}(-1)^{k+1}\binom{6}{k}(1-\frac{k}{6})^r$$

So, $P(X \le r) = 1 - \sum_{k=1}^{6}(-1)^{k+1}\binom{6}{k}(1-\frac{k}{6})^r$. The pmf then can be written as follows:

$$P(X) = \begin{cases} 0 \quad x < 6 \\ P(X \le x+1) - P(X \le x) \quad x \ge 6 \\ \text{(Using equation for } P(X \le r) \text{ above.)} \end{cases}$$

Note that it is not enough to count only outcomes where we get the sixth face in the r^{th} throw since this outcome remains valid as a positive event even when you are looking at more than r throws. From a counting viewpoint, of all the n^r outcomes after r throws we have to count all those r-tuples that have all six faces.

(c) A and B play a game where each guesses whether the BSE (Bombay Stock Exchange) index has risen or fallen at the end of the trading day (compared to start of day). Both put Rs.100/- in a bag at the start of the day. The person guessing correctly gets the full amount in the bag. If more than one guesses correctly then the money is equally shared. If there is no correct guess the money is carried over to the next day.

A observes that B's guesses are random (i.e. fair coin toss based). He tells B that his sister would also like to join in the game. A will put in Rs.200/- in the bag at start of day and tell both his and his sister's guesses at the end of the day. A's sister's guess is always the opposite of A's guess. What is the expected profit/loss made by A and his sister per day?

Solution:
(A+sister) always have one correct answer. So, when B is correct (assume half the time) A+sister will share the Rs.300/- with B. When B is wrong the entire Rs.300/- will go to A+sister. So, the expected value is: $150 \times \frac{1}{2} + 300 \times \frac{1}{2} = 225$. Since A+sister put Rs.200/- in the bag their expected profit is Rs.25/day.

[5,5,5=15]

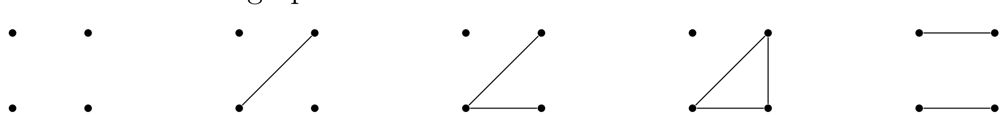
3. In this question $G = (V, E)$ is a simple graph. Remember a simple graph is one that does not have self loops or multi-edges.

(a) If $|V| = n$ how many simple graphs can you have in all (i.e. isomorphic graphs are counted separately)?

Solution:
In a simple graph between any two vertices there is one edge or no edge. There are ${}^nC_2$ pairs and for each pair we have two possibilities (edge present or absent) so the total number of simple graphs is: $2^{({}^nC_2)}$.

(b) For $|V| = 4$ draw the distinct (i.e. non-isomorphic) disconnected (i.e. have more than one component) simple graphs that are possible.

Solution:
There are 5 such graphs.



(c) A **clique** in a graph is a sub-graph where all vertices of the sub-graph are pairwise adjacent. So, if the clique has m vertices then the clique is K_m - a complete graph of order m. An

independent set in a graph is a set of vertices that are pairwise non-adjacent.

Argue that if $|V| = 6$ then G has a clique of size at least 3 or an independent set of size at least 3.

Solution:

This problem is identical to the one discussed in class that said that in a group of 6 persons there are at least 3 strangers or at least 3 persons who are mutually acquainted.

Choose some $v \in G$. It has 5 other vertices to which it can be adjacent. If v is adjacent to 3 or more vertices then if any of these neighbours of v are adjacent we have clique of size 3 by including v. If none of these vertices are adjacent then we have an independent set of size 3.

If v is adjacent to 2 or fewer vertces then consider the set of vertices not adjacent to v - this has size at least 3. If any two vertices in this set are not adjacent then together with v they will form an independent set of size 3. Alternately, if all of them are adjacent to each other then it will be a clique of size at least 3.

(d) Argue that if $\delta(G) \geq \frac{(n-1)}{2}$ where $|V| = n$ then G is connected. Using $K_{\lfloor n/2 \rfloor}$ and $K_{\lceil n/2 \rceil}$ as components of G show that the previous bound on $\delta(G)$ is tight.

Solution:

We will show that any two non-adjacent vertices u, v have a common neighbour. This implies that u and v are connected and consequently G is connected.

Let $N(v)$ denote the vertices that are adjacent to v. Since G is simple $|N(v)| \geq \delta \geq \frac{(n-1)}{2}$. Same is true for u. Since v and u are not adjacent $|N(v) \cup N(u)| \leq n - 2$.

We have $|N(v) \cap N(u)| = |N(v)| + |N(u)| - |N(v) \cup N(u)| \geq \frac{(n-1)}{2} + \frac{(n-1)}{2} - (n-2) = 1$. So, v, u have at least one common neighbour.

Let G be a graph containing $K_{\lfloor n/2 \rfloor}$ and $K_{\lceil n/2 \rceil}$ as components. Clearly, $\delta(G) = \lfloor n/2 \rfloor - 1$ and G is disconnected. So, the bound on $\delta(G)$ is tight.

[3,5,5,(5,2)=20]

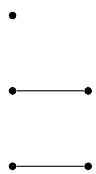
4. (a) Is 33333221 the degree sequence of some simple graph? Show details of your calculation.

Solution:

We carry out the recursive reduction given by the graphic sequence theorem.

33333221→2223221→3222221→111221→221111→10111→11110

The last sequence is easily seen to be graphic - see graph below.



So the initial degree sequence 33333221 is graphic.

(b) Show that a tree with maximum degree $\Delta > 1$ has at least Δ vertices of degree 1. Also, argue that this is a tight lower bound by showing how one can construct trees with n vertices and exactly Δ vertices of degree 1 satisfying $n > \Delta \geq 2$.

Solution:

There are multiple proofs possible. Note that a vertex with degree 1 is a leaf and we will refer to it as such below.

Proof 1: Let v be a node with degree Δ. Remove v. This will lead to Δ component trees that are sub-trees of the original tree. If a component tree is a single node then it will be a leaf of the original tree. If not, since it is a tree it will have at least one leaf node which will also be a leaf of the original tree. Thus the total number of leaves of the original tree is at least Δ.

Proof 2: We know that the sum of all degrees of the vertices of the tree is $2|E|$, that is, $2(n-1) = 2n-2$. Let us count this another way. Let l be the number of leaves in the tree. Every vertex other than the one with degree Δ and the leaves will have a degree of at least 2. So, the sum of all the degrees, $2n-2$, will be greater than or equal to $\Delta + l + 2(n-l-1)$. So, $2n-2 \geq \Delta + l + 2(n-l-1)$. This simplifies to $l \geq \Delta$.

To, see that the bound is tight consider the graph $K_{1,\Delta}$. This clearly has Δ leaves by construction. Each leaf node can be extended as a path by adding any number of nodes one after another. This increases n but the number of leaves remain unchanged at Δ. Clearly, this is possible for any choice of $\Delta \geq 2$ and n. Thus the bound is tight.

[5,(3,2)=10]

5. (a) You are given a $m \times n$ matrix template which is empty. You have to fill it with 4 distinct values such that no two values occur in the same row or same column. How many distinguishable matrices are possible?

Solution:

Let (r1,c1), (r2,c2), (r3,c3), (r4,c4) be the positions of the 4 values. We are want to find sequences (r1,r2,r3,r4,c1,c2,c3,c4) such that all the r_s and c_s are distinct. Clearly, we can choose r1 in m ways, r2 in (m-1) ways, r3 in (m-2) ways and r4 in (m-3) ways. Similarly for c1, c2, c3, c4 we get n, (n-1), (n-2), (n-3) ways. Since rows and columns are independent by the product rule we have: $m \times n \times (m-1) \times (n-1) \times (m-2)(n-2) \times (m-3)(n-3)$ ways such that no two values have the same row or column.

(b) A walking robot starts at the origin (0,0). In one second the robot can move one step east or one step west or one step north or one step south. The robot walks for 24 seconds, where its steps are equally divided between east, west, north and south. How many distinct walks can the robot make? (Your answer can be an expression).

> **Solution:**
>
> Here we get sequences of E, W, N, S, each occurring 6 times for a sequence of length 24. There is a bijection between the set of sequences and the set of walks. Clearly, the 6 occurrences of a particular letter (from E—W—N—S) in a sequence can be shuffled without changing the sequence. So, the total number of walks is: $\frac{24!}{6!6!6!6!} = \frac{24!}{6!^4}$.

(c) You have a sequence of n^2+1 distinct integers. Show that any such sequence will always have either an increasing or decreasing subsequence of length $(n+1)$. Also, show that n^2+1 is strict, in the sense that if we have n^2 integers then the property may not hold. (The subsequence need not consist of consecutive elements.)

> **Solution:**
>
> Let the sequence be $x_1, \ldots, x_{n^2+1}$. Let n_i be the number of terms in the longest increasing subsequence starting at x_i. If some $n_i \geq (n+1)$ we are done.
>
> We prove by contradiction. Assume all $n_i \leq n$. We have n^2+1 n_is to be distributed within n values. By the pigeon hole principle one of the values will contain at least $\lfloor \frac{(n^2+1)-1}{n} \rfloor + 1 = n+1$. That means at least $(n+1)$ values amongst the n_is are equal. We argue that the x_is associated with the above n_is form a decreasing subsequence. Let $n_i = n_j$ and $i < j$. Claim: $x_i > x_j$. Assume the contrary, i.e. $x_i < x_j$ (equality not possible since elements distinct). Then we have an increasing subsequence starting at x_i of length $n_j + 1$. So, $n_i \geq n_j + 1$ which is a contradiction.
>
> To see that n^2+1 is strict let $n=2$ and consider the sequence $3, 1, 4, 2$ of length 4. The longest increasing or decreasing subsequence is of length $2 < (n+1)$.

(d) Supposing we have 3 kinds of objects: k_1, k_2 and k_3. While the kinds are distinguishable, objects within a kind are not distinguishable. Assume we can pick 0, 1, or 2 objects of type k_1, 0 or 1 of type k_2 and 0 or 1 of type k_3. We want to calculate the number of distinguishable ways to pick k objects. If c_k represents this number find the generating function $G(x) = \sum_k c_k x^k$.

Generalize the above for the case when we have m types of objects $k_1, \ldots, k_m$ with $n_1, \ldots, n_m$ of each type and we want to calculate c_k the number of distinguishable ways to choose k objects where we are allowed to pick any number of objects of each type upto n_k, $k = 1..m$ for each type. In this case what is the generating function $G(x)$.

> **Solution:**
>
> We get finite power series as generating functions in both cases.
>
> We can pick from 0 to 2 of type k_1, 0 to 1 of type k_2, 0 to 1 of type k_3. Note

that within a type objects are indistinguishable but they are distinguishable between types. So, if k=3 then (1,1,1), (2,1,0) and (2,0,1) are all distinguishable, where the 3-tuple gives the number of objects of types k_1, k_2, k_3 respectively. The generating function can be written as (where we are using k_is as coeffcients):

$$\begin{aligned} G(x) &= [(k_1x)^0 + (k_1x)^1 + (k_1x)^2][(k_2x)^0 + (k_2x)^1][(k_3x)^0 + (k_3x)^1] \\ &= (1 + k_1x + k_1x^2)(1 + k_2x)(1 + k_3x) \\ &= 1 + (k_1 + k_2 + k_3)x + (k_1k_2 + k_2k_3 + k_1k_3 + k_1^2)x^2 + (k_1k_2k_3 + k_1^2k_2 + k_1^2k_3)x^3 \\ &\quad + k_1^2k_2k_3x^4 \end{aligned}$$

Note how the coefficient of x^k gives the ways in which k objects can be picked. Putting $k_1 = k_2 = k_3 = 1$ gives us the count of the number of ways k objects can be picked, that is c_k.

$$G(x) = 1 + 3x + 4x^2 + 3x^3 + x^4$$

By analogy the generalization is:

$$G(x) = (1 + x + x^2 + \ldots + x^{n_1})(1 + x + x^2 + \ldots + x^{n_2}) \ldots (1 + x + x^2 + \ldots + x^{n_m})$$

The coefficient c_k of x^k in $G(x)$ above will give a count of the number of ways in which k objects can be chosen.

[5,5,(6,4),(3,2)=25]

6. (a) An IITK student who wants to do a dual major in CSE must pass the following courses (ESc101 is a core course and done as part of the primary major):

CS201, CS202, CS203, CS210, CS220, CS251, CS252, CS330, CS335, CS340, CS345, DE1, DE2.

DE1, DE2 are chosen from the basket: {CS315, CS350, CS365, CS422, CS425, CS433, CS455}.

Course pre-requisites are shown below. $x < u$ means that course x is a pre-requisite for courses u. $<$ is the pre-requisite relation. Below, we are clubbing multiple relations in a single line. The meaning is obvious (the first line says ESc101 is a pre-requisite for CS201, CS210 and CS220).

ESc101 $<$ CS201, CS210, CS220
CS201 $<$ CS340, CS345
CS202 $<$ CS365, CS350
CS210 $<$ CS315, CS345, CS340, CS330, CS350, CS365
CS220 $<$ CS315, CS330, CS335, CS433, CS422
CS251 $<$ CS252
CS315 $<$ CS455
CS330 $<$ CS315, CS425, CS422

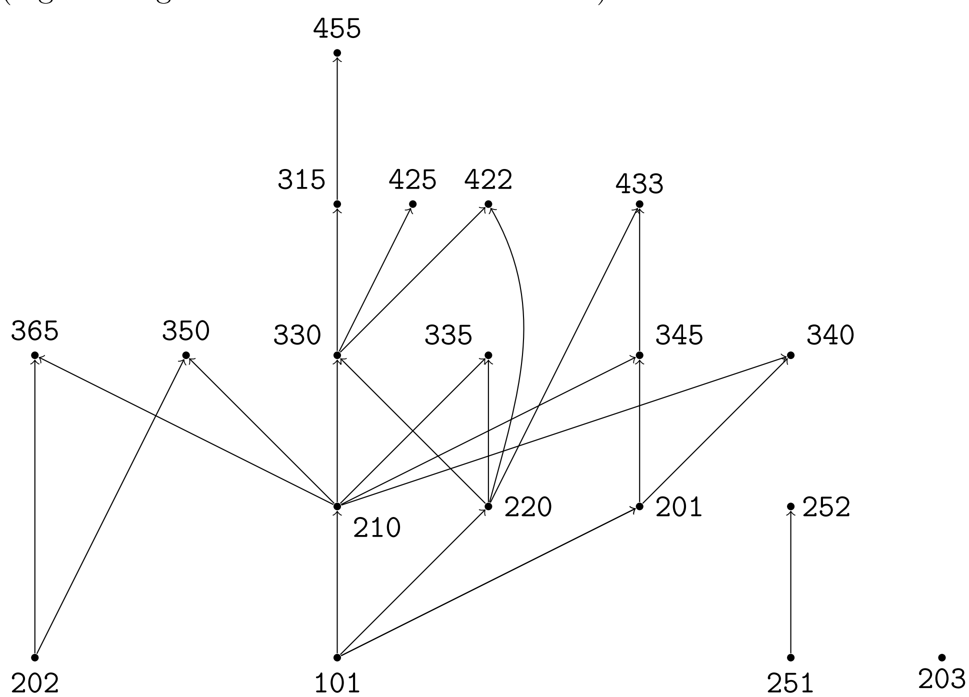
CS345 < CS433

i. Which of the properties: {reflexive, irreflexive, symmetric, antisymmetric, asymmetric, transitive} holds for the pre-requisite relation <.

> **Solution:**
> < is clearly irreflexive, asymmetric, transitive. Irreflexive and asymmetric implies anti-symmetric so it is also trivially anti-symmetric.

ii. Draw a DAG where a course is a vertex and there is a directed edge from x to v, if $x < v$. Draw only those edges that are necessary. Some edges will be implied based on the properties in part 6a)i), do not draw these edges.

> **Solution:**
> If the digraph is properly drawn it helps greatly with later parts of this question (a good diagram is worth a thousand words).



iii. Assuming all courses are offered in all semesters and a student can do at most 5 courses a semester what is the minimum number of regular semesters (ignore the summer semester) needed to complete the dual major requirement. Give a semester-wise list of courses. Show all calculations/reasoning.

> **Solution:**
> The digraph above immediately gives the answer. At least 4 semesters are required if we count the ESc101 semester as one semester. If we assume ESc101 was part of the core and do not count it then 3 semesters.

Many time tables are possible. Note that if we do 4 courses in semester 1 including ESc101 we can do only 4 courses in semester 2 and we have at least one course left over for semester 4. So, it is better to spread it out. Here is one possible break up:
Sem 1: ESc101, CS202, CS203 CS251
Sem 2: CS201, CS210, CS220, CS252
Sem 3: CS330, CS340, CS345, CS335
Sem 4: DE1, DE2
If we assume that ESc101 is done as part of the core and do not count it. We can do it in 3 semesters. Here is one possibility.
Sem 1: CS210, CS201, CS220, CS201, CS251
Sem 2: CS345, CS330, CS340, CS335, CS252
Sem 3: CS203, DE1, DE2

iv. If a student can do ESc101 and 4 other courses (in open elective slots) while completing the primary major what is the minimum number of extra semesters needed to complete the dual major requirement. Give a semester-wise list of courses. Show all calculations/reasoning.

Solution:
During primary major it is best to use the 4 OE slots for courses that are prerequisites. Assuming ESc101 is done before, one choice is: CS210, CS220, CS201, CS202. Then we complete the rest in 2 extra semesters.
Sem 1: CS330, CS335, CS340, CS345, CS251
Sem 2: CS252, CS203, DE1, DE2 (not all DEs are possible).

(b) S_1, S_2, S_3, S_4 are sets. Define:

$$X = (S_1 \cup S_2) \times (S_3 \cup S_4)$$
$$Y = (S_1 \times S_3) \cup (S_2 \times S_4)$$

i. Which of the following holds: $X \subset Y$, $X = Y$, $Y \subset X$? Briefly justify your answer.

Solution:
A simple cardinality argument shows that only $Y \subset X$ is possible. Let $S_i = \{i\},\ i = 1, 2, 3, 4$. Then $|X| = 4$ and $|Y| = 2$.
Other explicit counter examples that rule out $X = Y$ and $X \subset Y$ are also easy to construct.

ii. Consider the following proof that $X = Y$.

The proof argues that $\forall a, b,\ (a,b) \in Y \Leftrightarrow (a,b) \in X$

$$\begin{aligned}
(a,b) &\in Y && (1)\\
&\Leftrightarrow (a,b) \in (S_1 \times S_3) \cup (S_2 \times S_4) && (2)\\
&\Leftrightarrow (a,b) \in (S_1 \times S_3) \text{ or } (S_2 \times S_4) && (3)\\
&\Leftrightarrow (a \in S_1 \text{ and } b \in S_3) \text{ or else } (a \in S_2 \text{ and } b \in S_4) && (4)\\
&\Leftrightarrow (\text{either } a \in S_1 \text{ or } a \in S_2) \text{ and } (\text{either } b \in S_3 \text{ or } b \in S_4) && (5)\\
&\Leftrightarrow a \in (S_1 \cup S_2) \text{ and } b \in (S_3 \cup S_4) && (6)\\
&\Leftrightarrow (a,b) \in X && (7)
\end{aligned}$$

Where is the error in the proof? Correct it so that it becomes a proof of the claim you make in part 6b)i).

Solution:

There are multiple mistakes.

Since S_is are arbitrary sets it is possible $S_1 \cap S_2 \neq \emptyset$ and $S_3 \cap S_4 \neq \emptyset$. This implies that 'or else' in (4) and 'either or' in (5) should be replaced by plain 'or' which accounts for all 3 possibilities - $(a,b) \in S_1 \times S_3$, $(a,b) \in S_2 \times S_4$ or exists in both. This gives the alternates $(4')$ and $(5')$ respectively.

$$(a \in S_1 \text{ and } b \in S_3) \text{ or } (a \in S_2 \text{ and } b \in S_4) \quad (4')$$

$$(a \in S_1 \text{ or } a \in S_2) \text{ and } (b \in S_3 \text{ or } b \in S_4) \quad (5')$$

Now it is clear that while $(4' \Rightarrow 5')$ holds the converse, that is, $(5' \Rightarrow 4')$ does not because $(5')$ permits pairs from $S_1 \times S_4$ and $S_2 \times S_3$ which $(4')$ does not.

So, the correct proof has $(4') \Rightarrow (5')$. This breaks the chain of bi-implications and only permits $Y \subset X$.

[(2,6,5,4),(3,5)=25]