SIT718 Real world Analytics

Trimester 1, 2018

PRACTICE EXAM PAPER 1, Paper 2 SOLUTIONS

Question 1. Aggregation functions - calculations

(i) We can only determine PM(12,13,20,19) if we know p. If p=1 then translation invariance will mean that our output should be 16 (the arithmetic mean), however if p is not equal to 1 then this will not hold. Furthermore, in this case, we can tell that p is 1 since the arithmetic mean of 9,10,17 and 16 is 13.
On the other hand, since the power mean is homogeneous for all p, we can deduce that the result should be 26 for PM(18,20,34,32) [double 13 since all the inputs are doubled].

(a) Calculating for each function,

i) WAM

$0.3(0.1) + 0.6(0.9) + 0.05(0.3) + 0.05(0.2) = 0.03 + 0.54 + 0.015 + 0.01 = 0.595$

ii. PM

$(0.2(0.7^{-2}) + 0.25(0.2^{-2}) + 0.55(0.5^{-2}))^{(-1/2)} = 0.33599$ (to 5 decimal places)

iii. GM

$0.1^{0.5} \times 0.2^{0.3} \times 0.5^{0.1} \times 0.2^{0.1} = 0.1550$ (to 4 decimal places)

iv. OWA

$0.9\frac{1}{16} + 0.3\frac{3}{16} + 0.2\frac{5}{16} + 0.1\frac{7}{16} = 0.21875$

(b) By weights, x_4 is more important, however a power mean with $p = 3$ is affected more by high inputs (so likely to be more sensitive to changes in x_2. To decide, we need to check.

PM(0.67, 0.5, 0.39, 0.2) = 0.4037 (to 4 dp)

PM(0.67, 0.5, 0.39, 0.21) = 0.4053 (to 4 dp)

PM(0.67, 0.51, 0.39, 0.2) = 0.4068 (to 4 dp)

So it seems like increasing x_2 is better.

Question 2. **Interpreting parameters and weights**

(a)

i. $\mathbf{w} = \langle 0.1, 0.7, 0.2 \rangle$ suggests that the most important criterion is proximity to high achieving schools (x_2 scores), then proximity to public transport and price are much less important. If using a weighted arithmetic mean, then high scores, to an extent, can compensate for low scores, however it would be very hard for a house to score highly if it is not close to high achieving schools.

ii. $\mathbf{w} = \langle 0.4, 0.4, 0.2 \rangle$ suggests that Price (x_1) and Schools (x_2) are equally important and more important than being close to public transport. However the value of $p = -2$ means that the function will tend toward lower inputs. This means that a low score in proximity to public transport can have a high influence on the function and that all criteria should be reasonably satisfied if we are to obtain a high score.

iii. $\mathbf{w} = \langle 0.6, 0.2, 0.2 \rangle$ suggests that most of the criteria should be satisfied in order to obtain a high score. Most of the weight is allocated to the lowest input. For example, scores of 0, 1, 1 would only obtain an overall score of 0.4 (even though two of the criteria are perfectly satisfied).

iv.Looking at the singletons, the most important criterion is Price. The importance of proximity to schools and public transport are reasonably high alone, but nothing is gained when they are together. This means that these two criteria are more or less redundant (it's enough to satisfy one of them). If only the price were satisfied (e.g. an input of $\langle 1, 0, 0 \rangle$ then we would still obtain a score of 0.8.

(b) Using the OWA,

$$\text{OWA}(0.8,\ 0.2,\ 0.5) = 0.6(0.2) + 0.2(0.5) + 0.2(0.8) = 0.12 + 0.1 + 0.16 = 0.38$$

$$\text{OWA}(0.6,\ 0.5,\ 0.4) = 0.6(0.4) + 0.2(0.5) + 0.2(0.6) = 0.24 + 0.1 + 0.12 = 0.46$$

According to the OWA model, house b seems better. This is because even though house a has a very good price, it is not very close to good schools. House b has better scores overall, with no criterion falling below 0.4. However both scores are still less than 0.5.

Question 3 Paper 2 **Transformations**

(a) All variables should be increasing with the output, i.e. better performance in terms of Year 12 results, Year 9 test results, student progress and student support should give the school a better overall result. We therefore don't need to use any kind of negation or decreasing function.

For criterion 1,
Assuming the percentile scores are given between 0 and 100, we can use

$$x' = \frac{x}{100}$$

(it is not necessary to subtract the minimum)

For criterion 2,
Standardised testing can be roughly transformed to the unit interval using

$$x' = 0.15x + 0.5$$

This means approximately 99.7% of the data should lie between 0.05 and 0.95. Alternatively, denote the minimum school's score by b and the best school's by a, we can use

$$x' = \frac{x-b}{a-b}$$

For criterion 3,
These scores are already between 0 and 1. No transformation required, however since the distribution is skewed toward high values we could use $x' = x^2$ in order to distribute the values a little more evenly.

For criterion 4,
The highest score is 10 and the lowest score (from the histogram is 4. We can use

$$x' = \frac{x-4}{6}$$

(b) I would place slightly higher importance on Student support and student progress, since these scores indicate a school's attempt to help the lowest performing students. I would give year 12 results the lowest weight, since this is largely influenced by student ability on entering year 12. (* note, opinions expressed here can be subjective, they just need to correspond with the weights you choose).
Therefore, the weighting vector I would use (with, e.g. a weighted power mean) would be

$$\mathbf{w} = \langle 0.15, 0.25, 0.3, 0.3 \rangle$$

Question 3 Paper 1. **Data Analysis**

(a) The AM values give a baseline performance that gives as an idea of the improvement made by using customised aggregation functions.

The extent of improvement is reasonable, but not great. For outputs given over the unit interval (0 to 1), the average error is in the vicinity 0.15 for most functions, which means we're at least 10% or so out (on average).

The fact that the PM with $p = 2$ performs worse means that the data probably does not tend toward higher inputs (i.e. we need more of the criteria satisfied in order to get a high output - or in the context of the question, that just having a gym in an area with residents in the right age bracket (or the right personal income etc) is not enough to ensure a high number of gym users). (This is supported by the fact that the OWA function that is fit to the data has higher weight allocated to the lower inputs.)

The Choquet integral is the best performing function, however this is not surprising given that the Choquet integral generalises the OWA and WAM and is more flexible.

(b) i . From the WAM and PM, i t seems l ike the most i mportant variable i s the personal income variable - meaning that this is the most influential in determining whether a gym is popular or not. However looking at the values of the Choquet integral, this variable seems less important, unless in the presence of variable 3. In fact, with the Choquet integral, any of the variables alone are not too important but the values for the pairs (13) and (23) are quite high.

ii. As discussed previously, the values tend toward lower values. This is suggested by the OWA, i.e. the orness value is

$$0.5 \times (0.396) + 1 \times (0.214) = 0.412$$

We also note again that the power mean that tended toward high values did not fit very well. The Choquet integral with low values for singletons also suggests that lower values are favoured.

iii. Based on the pairs in the fuzzy measure of the Choquet integral, there is a superadditive/complementary effect between the personal income and local fitness centres variables and also between the age and fitness centres variables. On the other hand, there is not much interaction between the age and personal income variables.

uestion 4 Paper 1 *Linear Programming with Two Variables and Graphical Method*

Let x_1 be the amount of radiation (in Gy) to be delivery from Beam 1, and x_2 be the amount of radiation (in Gy) to be delivery from Beam 2. The LP model is as below.

The objective is to minimize radiation delivered to the healthy anatomy:

$$\min z = 0.4x_1 + 0.5x_2.$$

Subject to constraints:

1. Radiation delivered to critical tissues cannot exceed 27 Gy:

$$0.3x_1 + 0.1x_2 \leq 27$$

2. Radiation delivered to entire tumour region at least 60 Gy:

$$0.5x_1 + 0.5x_2 \geq 60$$

3. Radiation delivered to centre of tumour at least 70 Gy:

$$0.7x_1 + 0.35x_2 \geq 70$$

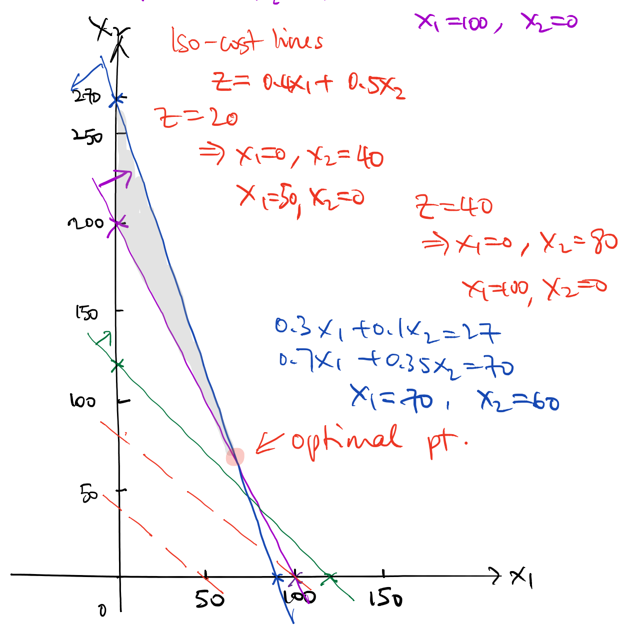
4. Bound constraints:

$$x_1, x_2 \geq 0$$

Q1 $0.3x_1 + 0.1x_2 = 27 \Rightarrow x_1=0,\ x_2=270$; $x_1=90,\ x_2=0$

$0.5x_1 + 0.5x_2 = 60 \Rightarrow x_1=0,\ x_2=120$; $x_1=120,\ x_2=0$

$0.7x_1 + 0.35x_2 = 70 \Rightarrow x_1=0,\ x_2=200$; $x_1=100,\ x_2=0$



uestion 4 Paper 2 Objective function: to minimize cost of hiring ground agents

$$\min z = 170x_1 + 160x_2 + 175x_3 + 180x_4 + 195x_5$$

Subject to time period constraints: (top row 6:00-8:00, in same order as presented in the table)

$$
\begin{aligned}
x_1 &\geq 48 && (1)\\
x_1 + x_2 &\geq 79 && (2)\\
x_1 + x_2 &\geq 65 && (3)\\
x_1 + x_2 + x_3 &\geq 87 && (4)\\
x_2 + x_3 &\geq 64 && (5)\\
x_3 + x_4 &\geq 73 && (6)\\
x_3 + x_4 &\geq 82 && (7)\\
x_4 &\geq 43 && (8)\\
x_4 + x_5 &\geq 52 && (9)\\
x_5 &\geq 15 && (10)\\
x_1, \ldots, x_5 &\geq 0 && (11)
\end{aligned}
$$

Question 5 Game theory, Shortest Path, critical path Problems

Note: LP models of games with no saddle point are within the scope. Critical path problems are within the scope of the exam as well, even though this practice exam paper does not have a question on this.

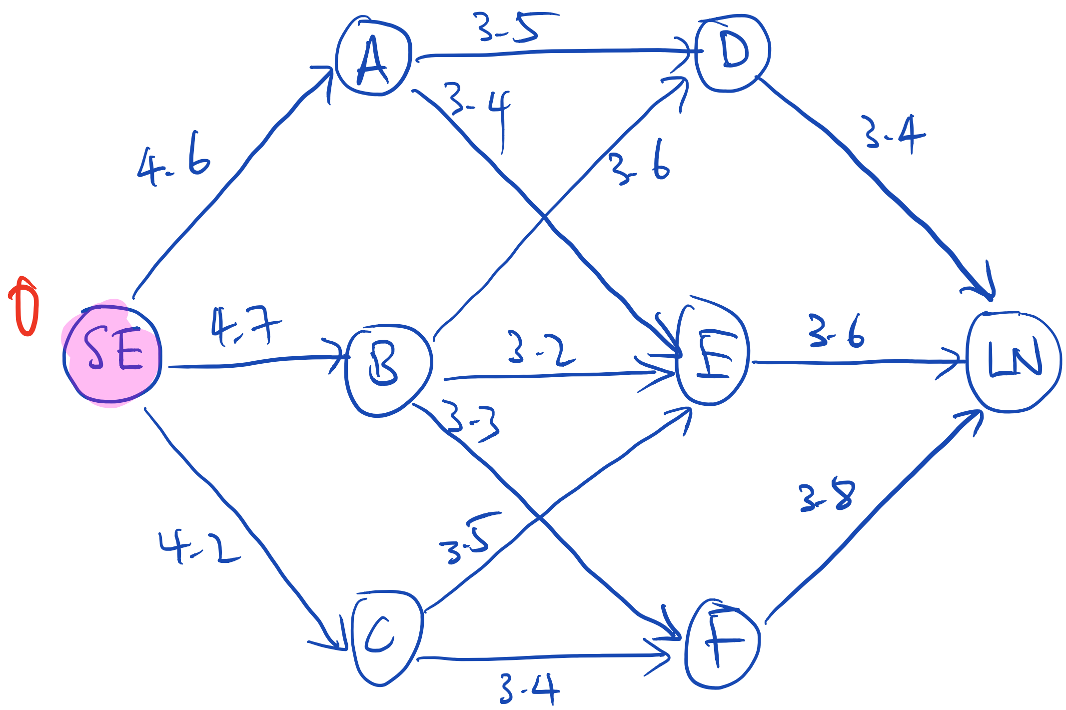
(i)

	A_1	A_2	A_3	A_4	Row Min
a_1	5	2	1	3	1
a_2	3	3	-2	4	-2
a_3	-1	3	-1	-3	-3
a_4	2	-4	1	-2	-4
Col Max	5	3	1	4	

Since min col max = max row min = 1, the game has an optimal pure strategy at strategy pair (a_1, A_3).

(ii)

see below:

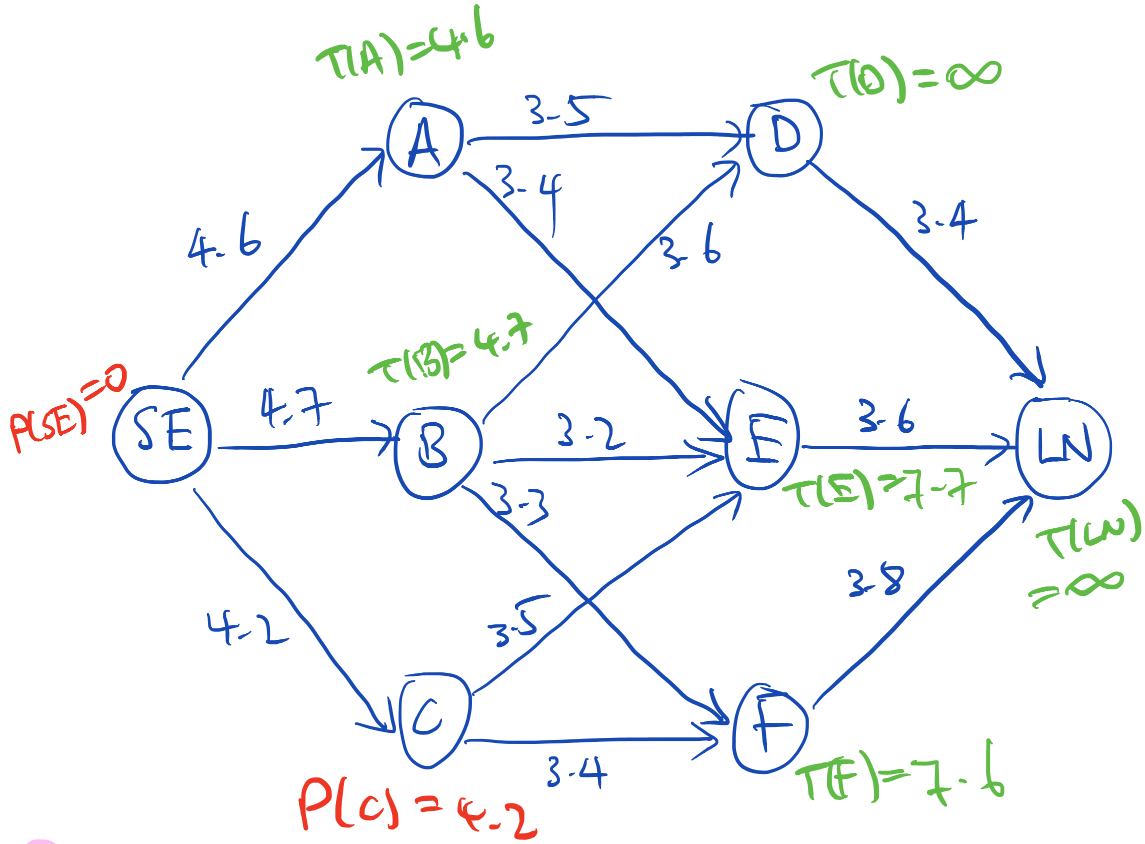


Round 1 $P(SE) = 0$, $\mathcal{L} = \{SE\}$

$T(A) = 4.6$, $T(B) = 4.7$, $T(C) = 4.2$,

$T(D) = \infty$, $T(E) = \infty$, $T(F) = \infty$

$T(LN) = \infty$



Round 2 $\mathcal{L}=\{SE, C\}$ $P(SE)=0$, $P(C)=4.2$

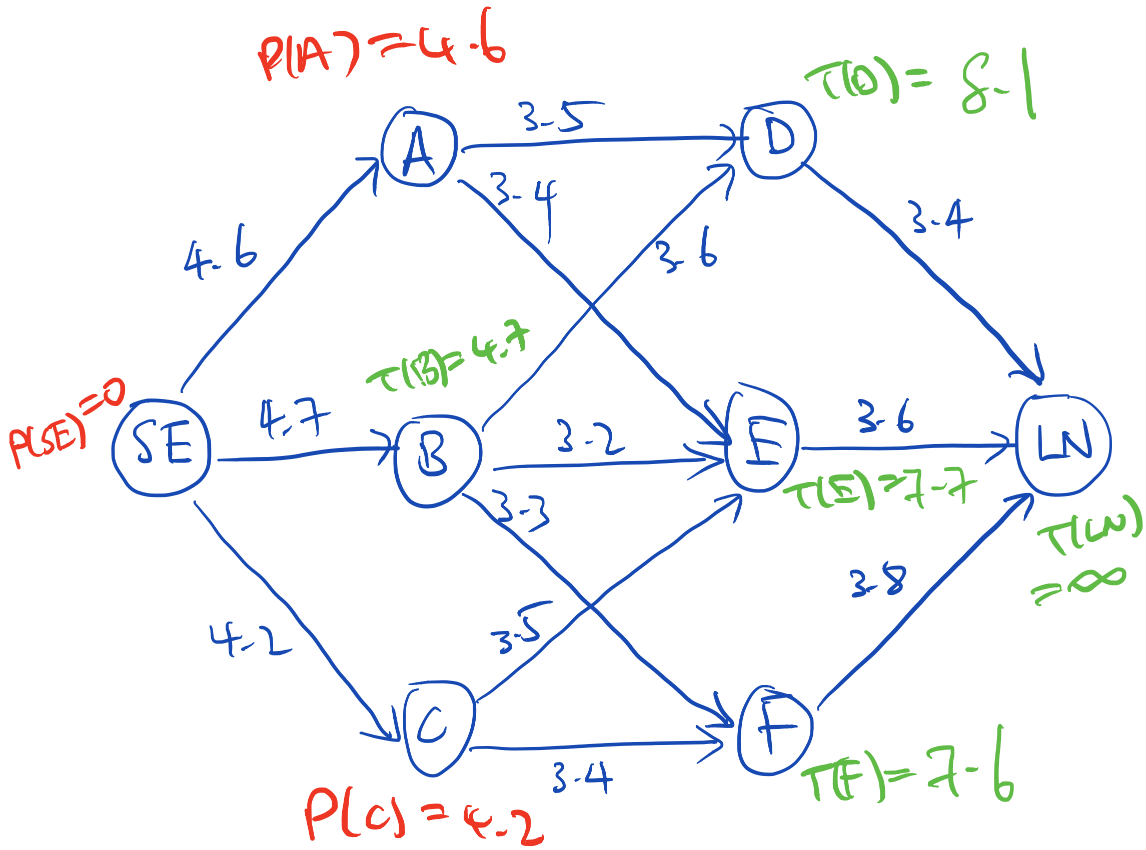
$T(A)=4.6$, $T(B)=4.7$

$T(D)=\infty$

$T(E)=\min\{\infty, 4.2+3.5\}=7.7$

$T(F)=\min\{\infty, 4.2+3.4\}=7.6$

$T(LN)=\infty$



Round 3 $\quad \mathcal{L} = \{SE, C, A\}$ $P(SE) = 0$, $P(C) = 4.2$

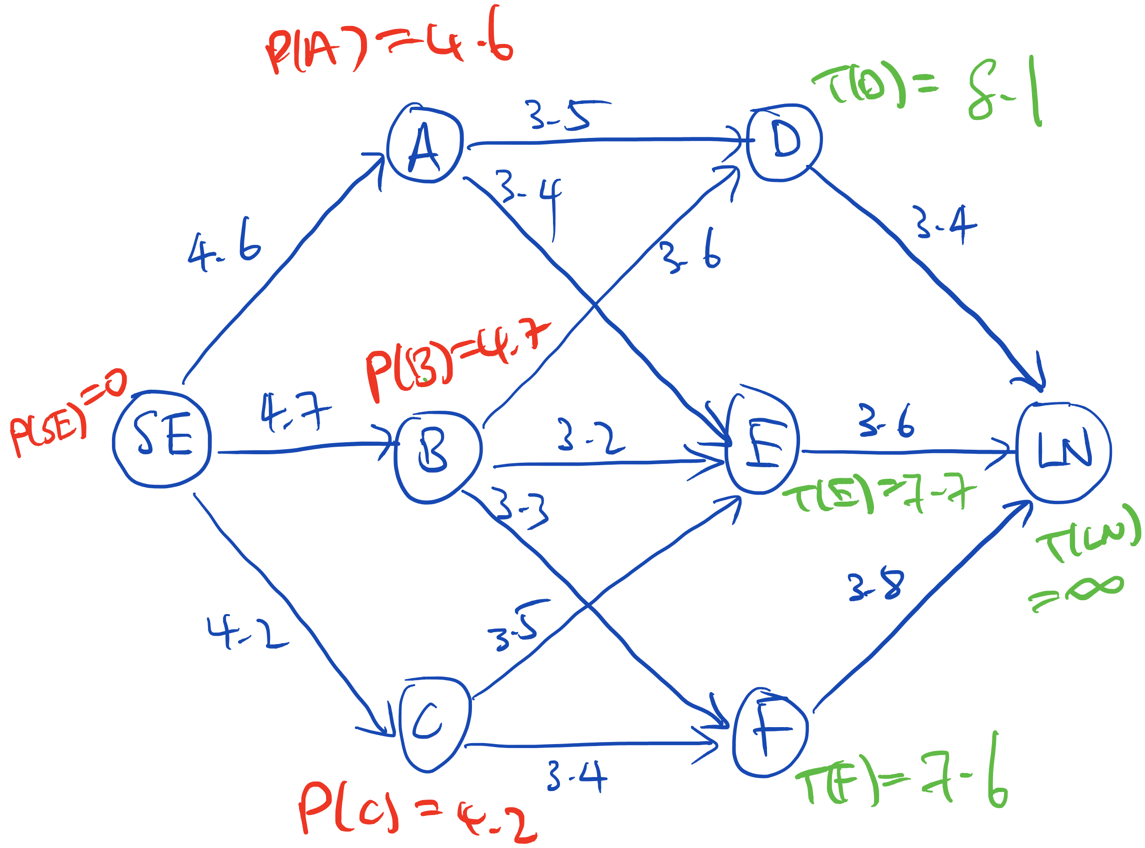
$P(A) = 4.6$, $T(B) = 4.7$

$T(D) = \min\{\infty, 4.6 + 3.5\} = 8.1$

$T(E) = \min\{7.7, 4.6 + 3.4\} = 7.7$

$T(F) = \min\{\infty, 4.2 + 3.4\} = 7.6$

$T(LN) = \infty$



Round 4 $\mathcal{L}=\{SE, C, A, B\}$ $P(SE)=0$, $P(C)=4.2$

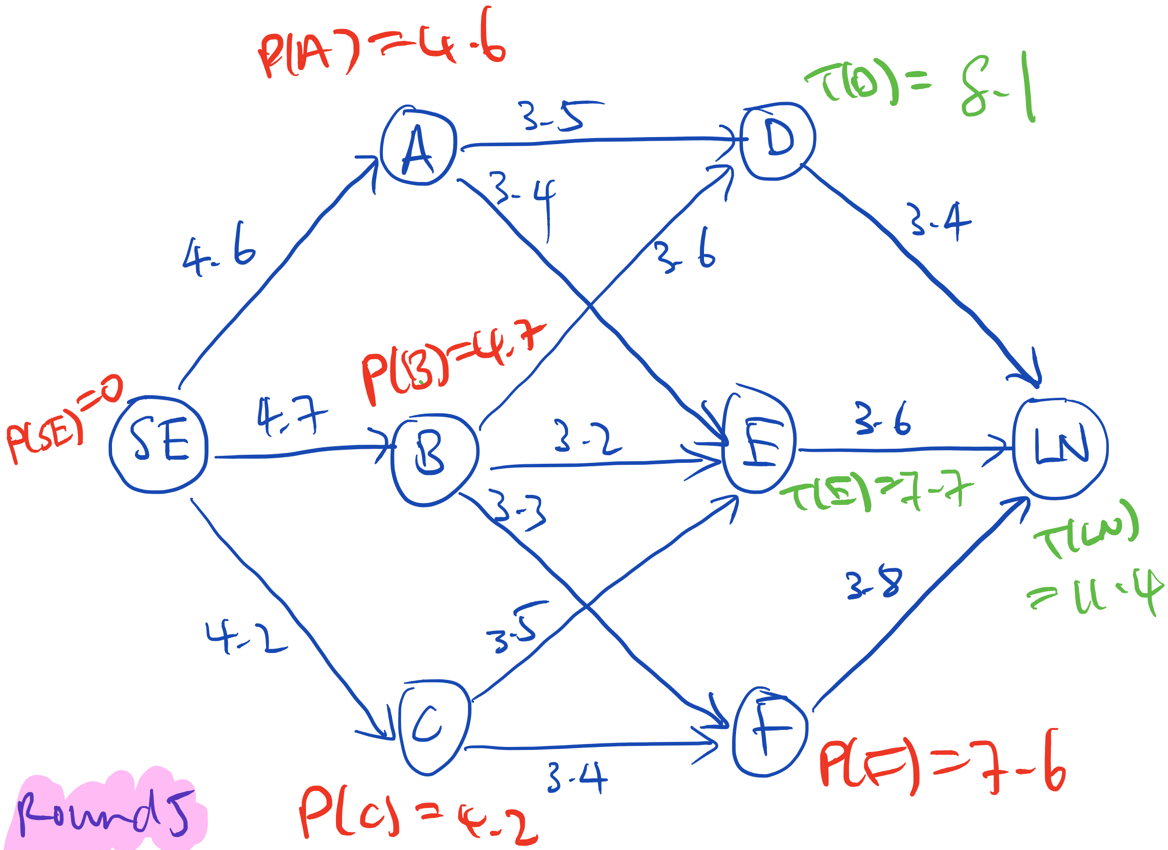
$P(A)=4.6$, $P(B)=4.7$

$T(D) = \min\{8.1,\ 4.7+3.6\} = 8.1$

$T(E) = \min\{7.7,\ 4.7+3.2\} = 7.7$

$T(F) = \min\{7.6,\ 4.7+3.3\} = 7.6$

$T(LN) = \infty$



Round 5

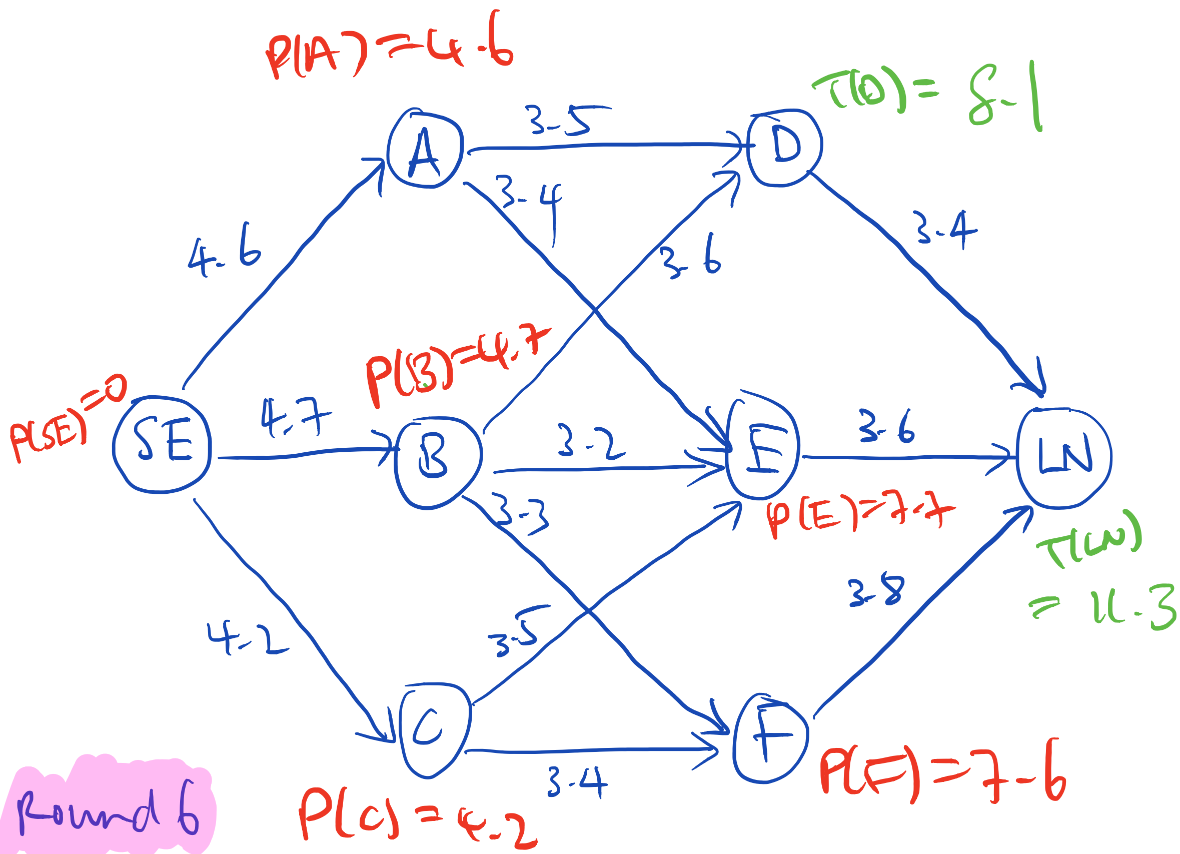
$L = \{SE, C, A, B, F\}$ $\quad P(SE) = 0, \; P(C) = 4.2$

$P(A) = 4.6, \quad P(B) = 4.7$

$T(D) = \min\{8.1, \; 4.7 + 3.6\} = 8.1$

$T(E) = \min\{7.7, \; 4.7 + 3.2\} = 7.7$

$T(LN) = \min\{\infty, \; 7.6 + 3.8\}$
$= 11.4$



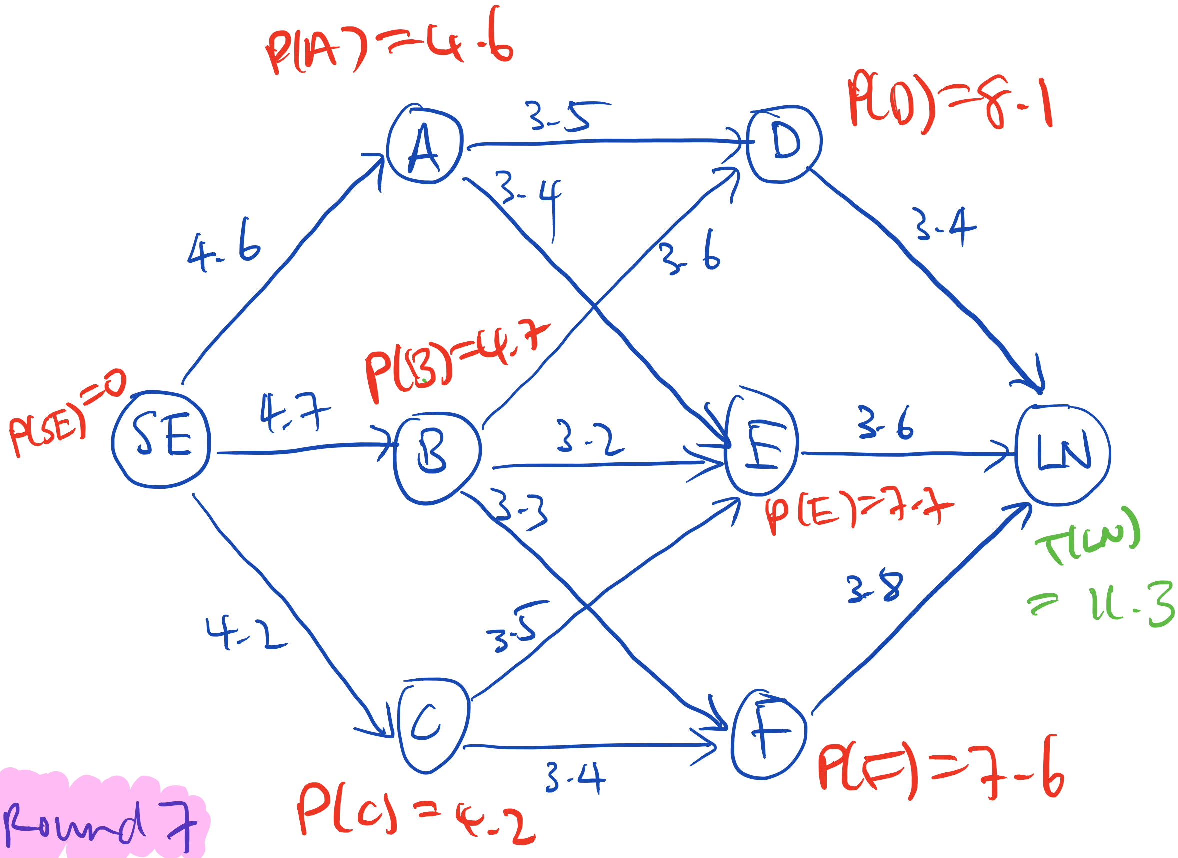
Round 6

$\mathcal{L} = \{SE, C, A, B, F, E\}$ $\quad P(SE) = 0,\ P(C) = 4.2$

$P(A) = 4.6,\quad P(B) = 4.7,\ P(E) = 7.7$

$T(D) = \min\{8.1,\ 4.7 + 3.6\} = 8.1$

$T(LN) = \min\{11.4,\ 7.7 + 3.6\}$
$= 11.3$



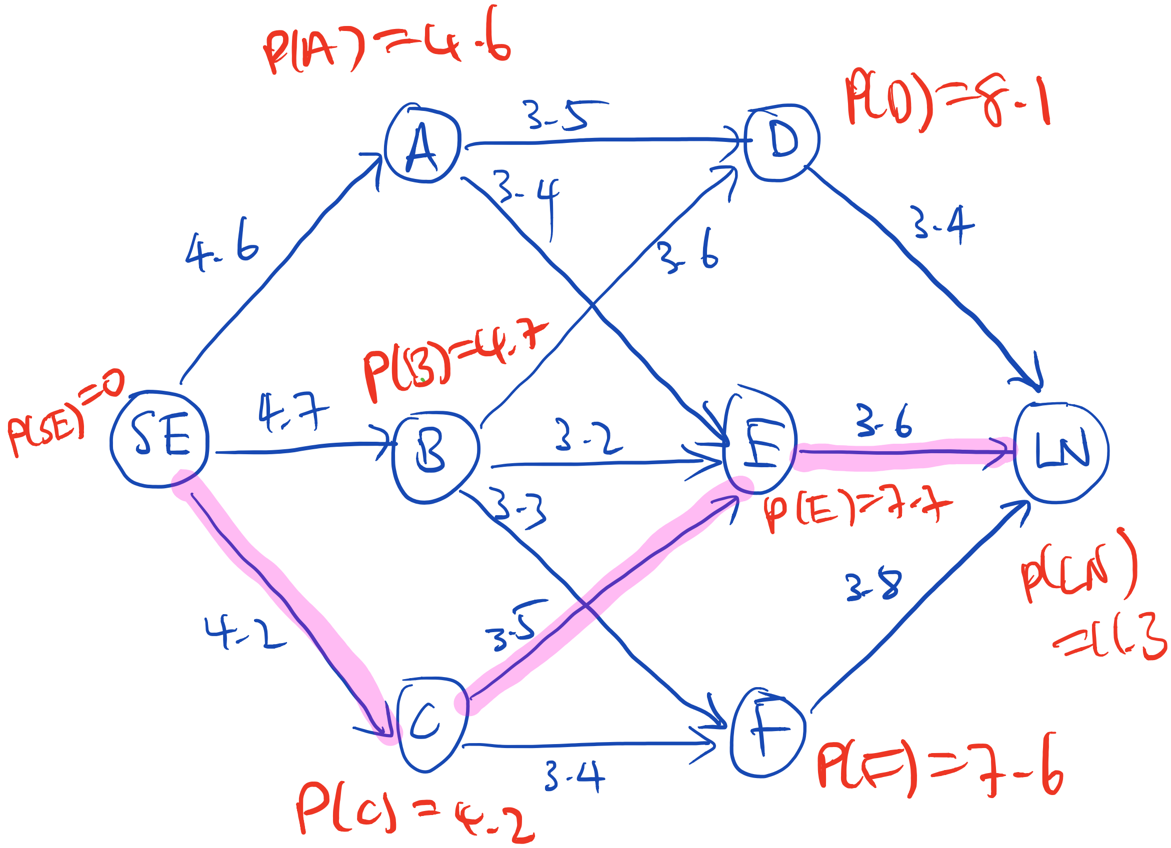
Round 7

$\mathcal{L} = \{SE, C, A, B, F, E, D\}$ $P(SE)=0$, $P(C)=4.2$

$P(A)=4.6$, $P(B)=4.7$, $P(E)=7.7$

$P(D)=8.1$

$$T(LN) = \min\{11.3,\ 8.1+3.4\}$$
$$= 11.3$$



Round 8   P(LN) = 11.3

Backtrack $P(j) - P(i) = c_{ij}$

SE → C → E → LN

Question 6 *[20 marks]Transportation Problems*

Let x_{ij} be a decision variable, for $i = 1, 2, 3$, with 1–Mitchellstown, 2–Whiteheads Creek, and 3–Locksley, and $j = 1, 2, 3, 4$, with 1–Melbourne, 2–Sydney, 3–Brisbane, 4–Adelaide.

The objective function is given by:

$$\begin{aligned}\min z = \quad & 464x_{11} + 513x_{12} + 654x_{13} + 867x_{14} \\ & +352x_{21} + 416x_{22} + 690x_{23} + 791x_{24} \\ & +995x_{31} + 682x_{32} + 388x_{33} + 685x_{34}\end{aligned}$$

Subject to Supply constraints:

$$\begin{aligned}x_{11} + x_{12} + x_{13} + x_{14} &\leq 75 \\ x_{21} + x_{22} + x_{23} + x_{24} &\leq 125 \\ x_{31} + x_{32} + x_{33} + x_{34} &\leq 100\end{aligned}$$

and demand constraints:

$$\begin{aligned}x_{11} + x_{21} + x_{31} &\geq 80 \\ x_{12} + x_{22} + x_{32} &\geq 65 \\ x_{13} + x_{23} + x_{33} &\geq 70 \\ x_{14} + x_{24} + x_{34} &\geq 85\end{aligned}$$

all $x_{ij} \geq 0$

[Note: Transportation problems with shortages or surplus are within the scope of the exam as well.]