ASSIGNMENT 1F – Due on Sunday 11:55pm of Week 12.

TOPICS TESTED: SIMULATION & FURTHER MODELLING

Do's & Don'ts:

*** All answers in this assignment can be done in any colour **except for red colour** as this is the colour your tutor will use to mark your assignment. Thus, answers in red colour will not be graded (even correct ones) ***

*** If you choose to handwrite your answer (scan and submit it electronically), make sure your **handwriting is readable** – this is a good practice for your exam. Failure to comply will result in a lack of marks awarded if your writing is unreadable and there will be no exception for this. ***

*** **Each sub-assignment will be worth 2.5%** towards your total score. Sub-assignments will have different point distributions within them as we aim to focus your attention to the important areas; however, the score for a sub-assignment remains 2.5%. ***

*** These questions are meant for you to solve **independently**, we encourage students to figure out the questions themselves as it would be good for their understanding of the topics; however, please feel free to consult your tutors if needed. **Plagiarism** (either from using online sources or copying the answers from your classmates) will be punished accordingly. ***

*** As this is considered to be an assignment (albeit a sub assignment), requests for special consideration or extension must be submitted at least **2 days BEFORE THE DEADLINE.** The due date is on Sunday, so the latest day you can ask for extension is on Friday (the last official working day of the week for the teaching team). Please follow Monash guideline to request for extensions (medical certificates, doctor or GP letter, etc). Emergencies are to be adjusted individually. ***

*** **No R or any other programming languages should be used in solving these questions**. All work for this assignment needs to be done manually, less the use of **non-programmable calculator** (this also applies to your Final Exam). Tutors are not required to answer questions in the difference between manual calculation and programmed calculation***

*** **Late submission is 10% per day, after 5 days you will be given no marks**. Late submission is calculated as following: If you get 70% on this assignment and you are late for 2 days (you submit on Tuesday), your scores is now 70% -20% (2x10% per day) = 50%. This is done to ensure that the teaching team can release your result as soon as possible so that you can review on your mistakes and have a better study experience. ***

*** Please **show all working** in answering questions, your score will be **halved** if you don't comply***

*** Assignments shall be marked completely in **two weeks' time** according to Monash Policies. If there are any changes to the marking time, we will duly inform you. Solutions **will not be released** for this assignment; you can come to the tutorial and ask for an explanation about how to solve the questions after scores are released. ***

*** Please don't send emails to tutors asking for suggestions, we have Moodle and consultations for that, In writing your inquiries on Moodle please try to be clear in your problem and not revealing your working to others as this might be counted as plagiarism on your part. A good format for inquiry topic would be "Assignment 1a – Tutorial 10 (your tutorial slot) – Question about median "***

*** Assignments need to be submitted in **PDF** format. Failure to comply will result in 30% penalty***

*** Filename format for submitting assignment "Assignment1A_StudentId.pdf". File with wrong format incurs 30% penalty ***

QUESTIONS:

SIMULATION: (10 Marks Total)

1) Inverse Sampling (3 marks): Given the **PDF**: $q(x) = \frac{1}{2} * e^{-0.5x}$ with x > 0. Please do the inverse sampling process for this equation given that the random samples from Uniform(0,1) = {0.2, 0.4, 0.6, 0.8}. Show all working.

ANSWER:

1) Inverse Sampling:

Given:

$$PDF(x) = q(x) = \frac{1}{2} e^{\frac{-1}{2}x} \qquad x > 0$$

$$\Rightarrow [0 \le x \le \infty]$$

Solution:

Step1: calculating the CDF from PDF

$$CDF(x) = \int_{-\infty}^{x} f(t)\, dt$$

$$= \int_{0}^{x} \frac{1}{2} e^{\frac{-1}{2}t}\, dt$$

$$= \frac{1}{2} \int_{0}^{x} e^{\frac{-1}{2}t}\, dt$$

$$= \frac{1}{2} \left. \frac{e^{\frac{-1}{2}t}}{\left(\frac{-1}{2}\right)} \right|_{0}^{x}$$

$$= -\left[e^{\frac{-1}{2}t} \right]_{0}^{x}$$

$$= -\left[e^{\frac{-1}{2}x} - 1 \right]$$

$$= 1 - e^{\frac{-1}{2}x} \quad // \qquad (x > 0)$$

1

Step 2: Calculating the quantile Function.

Given that $CDF(x \geq 0)$

$$= 1 - e^{\frac{-1}{2}x} \quad \text{①}$$

Finding the Inverse:

$$CDF(x) = y \quad \text{②}$$

From ①, ②:

$$y = 1 - e^{\frac{-1}{2}x}$$

$$e^{\frac{-1}{2}x} = 1 - y$$

$$\frac{-1}{2}x = \ln(1-y)$$

$$x = -2\ln(1-y) \,//.$$

Step 3:

Random samples from:

$$\text{uniform}(0,1) = \{0.2, 0.4, 0.6, 0.8\}$$

DOING the INVERSE SAMPLING ON THE RANDOM samples we get:

$$[-2\ln(1-0.2), -2\ln(1-0.4), -2\ln(1-0.6), -2\ln(1-0.8)]$$

$$= [-2\ln(0.8), -2\ln(0.6), -2\ln(0.4), -2\ln(0.2)]$$

$$= [0.4463, 1.0216, 1.8326, 3.2189] \,//.$$

②

2) Rejection Sampling (4 marks):

Given the following **PDF** $f(x) = -\frac{3}{32}(x^2 + 4x)\ with - 4 < x < 0$

a. Write down the algorithm for Rejection Sampling? (**1 mark**)
b. Find the constant **C**, what is the percentage of rejection and acceptance here? (**1.5 mark**)
c. What is the percentage of rejection if you only want to sample for x within [-3.5, -0.5], write down the whole algorithm for this case? (**1.5 mark**)

ANSWER:

2) REJECTION SAMPLING :

Given : $PDF(x) = \frac{-3}{32}(x^2+4x)$

$(-4 < x < 0)$

a) Algorithm for Rejection Sampling :

1) Here, we first sample x as an uniform distribution in the range $(-4, 0)$. [This is our TARGET DISTRIBUTION] $\rightarrow f(x)$

2) Then, we sample U as an uniform distribution in the range $(0, 1)$ [This is our PROPOSED DISTRIBUTION]. $\rightarrow P_{prop}(x)$.

3) If $U \le \frac{C\, f(x)}{P_{prop}(x)}$

Accept x

Else

Reject x

RETURN TO STEP 1

4) Return x

Here : $P_{prop}(x)$: Proposed Distribution UNIFORM DISTRIBUTION from -4 to 0.

$\Rightarrow P_{prop}(x) = \frac{1}{(b-a)} = \frac{1}{0-(-4)} = \frac{1}{4}$

2

b) $c\, f(x) \leq P_{prop}(x)$

where $f(x) = PDF(x)$

$P_{prop}(x) =$ Proposed Function.

Now $P_{prop}(x) = \frac{1}{4}$ For $x \in (-4, 0)$

$$\Rightarrow \quad c\, f(x) \leq \frac{\frac{1}{4}}{f}$$

$$\Rightarrow \quad c = \frac{(1/4)}{f(\hat{x})}$$

where $f(\hat{x}) =$ Max. value of $f(x)$ in the RANGE: $(-4, 0)$

$= \frac{3}{8}$ at $x = -2$

$$\Rightarrow \quad c = \frac{1}{4 \times 3/8} = \frac{2}{3}$$

c

Hence, Acceptance Percentage $= c\% = \frac{2}{3} \times 100 = 66.67\%$

Rejection Percentage $= (1-c)\% = \frac{1}{3} \times 100 = 33.33\%$

Now,

Acceptance % $= c \underbrace{\int_{-4}^{0} f(x)\, dx}_{=1} \times 100\% = c \times 100\%$

$= \frac{2}{3} \times 100\% = 66.67\%$

lly, Rejection % $= 1 - c \int_{-4}^{0} f(x) dx \times 100\% = (1-c) \times 100\%$

$= (1/3) \times 100\% = 33.33\%$

c) GIVEN:

New sampling RANGE : $[-3.5; 0.5]$

Algorithm :

1. Here, we first sample x as an uniform distribution in the range $[-3.5, -0.5]$ [Target Distribution]
2. Then we sample U as an uniform distribution in the range $[0, 1)$
3. If $U \leq \frac{c\, f(x)}{P_{prop}(x)}$

   Accept x

   Else

   Reject x

   RETURN TO step 1.
4. Return x

Here : $P_{prop}(x)$ = Proposed distribution

uniform distribution from $[-3.5, -0.5]$

$$\Rightarrow P_{prop}(x) = \frac{1}{b-a} = \frac{1}{-0.5 - (-3.5)} = \frac{1}{3}$$

Now,

$$c\, f(x) \leq P_{prop}(x)$$

$$c \leq \frac{(1/3)}{\hat{f(x)}}$$

$\hat{f(x)}$ = Max value of $f(x)$ in the RANGE : $[-3.5, -0.5]$

$$c = \frac{1/3}{3/8} = \frac{8}{9} = \frac{3}{8} \text{ at } (x = -2)$$

3

Percentage of Rejection

$$= 1 - c\int_{-3.5}^{-0.5} f(x)\,dx$$

$$= 1 - \frac{8}{9}\int_{-3.5}^{-0.5} \frac{-3}{32}(x^2+4x)\,dx$$

$$= 1 - \frac{8}{9}\left(\frac{-3}{32}\right)\int_{-3.5}^{-0.5} (x^2+4x)\,dx$$

$$= 1 + \frac{1}{12}\underbrace{\int_{-3.5}^{-0.5} (x^2+4x)\,dx}$$

INTEGRAL VALUE :-

$$\int_{-3.5}^{-0.5} (x^2+4x)\,dx$$

$$= \left.\frac{x^3}{3}\right|_{-3.5}^{-0.5} + \left.\frac{4x^2}{2}\right|_{-3.5}^{-0.5}$$

$$= \frac{1}{3}\left[(-0.5)^3-(-3.5)^3\right] + 2\left[(-0.5)^2 - (-3.5)^2\right]$$

$$= 14.25 - 24$$

$$= -9.75$$

Now, $\Rightarrow 1 + \frac{1}{12}(-9.75)$

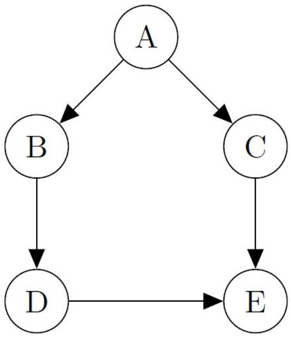
$\Rightarrow 0.1875$ //

HENCE,

% of REJECTION = $(1-C)$ %

= 18.75 % //

3) Gibbs Sampling (3 Mark)



a. Write a factorised probability expression for the full joint distribution $p(A, B, C, D, E)$. Please order the terms so that a sequential sampler could be used to sample each conditional probability distribution in turn reading from left to right. (**1 Mark**)
b. Which variables, if any, are independent of C when controlling for A i.e., assuming A is known (**1 Mark**)
c. If the network described above was conditioned on B = b, what four proportional conditional probability distribution functions would one need to be able to calculate in order to perform this type of sampling? Please write out how you would compute them. (**1 Mark**)

<u>ANSWER:</u>

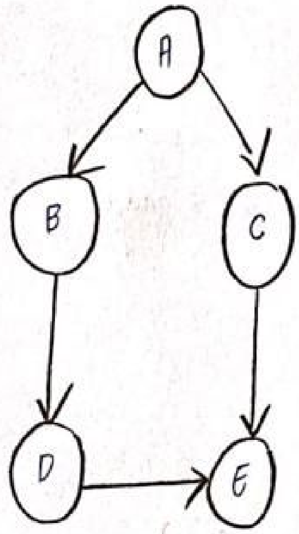
b) B is INDEPENDENT of C

Also, D is a CHILD of B so - D is also independent of C.

Hence, variables which are independent of C : B and D

(Other Answers, mentioned below)

3) GIBBS sampling :



a) Factorised probability expression for the full joint distribution

$= p(A, B, C, D, E)$

STEPS:

1) List each variable in the list : A, B, C, D, E

2) Associate every variable with it's parents in the graph

A( ), B(A), C(A), D(B), E(C, D)

3) we notice : A has NO PARENT
C, B have A as their parent
D has B as it's parent
E has C, D as it's parent

Lets Reorder the variables so that parents occur after they are first introduced as variables.

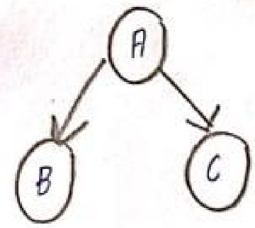
⇒ A( ), B(A), C(A), D(B), E(C, D)

4) Probability Formula :

$P(A, B, C, D, E) = p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|C, D)$ //

(4)

b)



→ Now, A is the COMMON CAUSE of B and C.

→ Here, A is the EVIDENCE.

→ Given that A : KNOWN

B is INDEPENT of C.

c) Given: Network was conditioned on $(B=b)$

Solution: In order to perform this sampling, we will need 4 conditional probability distribution functions:

we build a Gibbs sampler as follows:

1) Initialise A, C, D, E somehow.

2) sample A according to,

$$p(A \mid B,C,D,E) \propto p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|C,D)$$
$$\propto p(A) \cdot p(B|A) \cdot p(C|A)$$

3) sample C according to,

$$p(C \mid A,C,D,E) \propto p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|C,D)$$
$$\propto p(C|A) \cdot p(E|C,D)$$

4) sample D according to,

$$p(D \mid A,B,C,E) \propto p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|C,D)$$
$$\propto p(D|B) \cdot p(E|C,D)$$

5) sample E according to,

$$p(E \mid A,B,C,D) \propto p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|C,D)$$
$$\propto p(E|C,D)$$

6) Skip up to step 2 and continue.