Δ

CS 112 2$^{\text{nd}}$ Mid-term Spring 2017

4:00 am–5:50 pm   May 17$^{\text{th}}$

Notes:

1: The mid-term is closed book, closed notes.
2: Calculator allowed. No other electronic devices allowed.
3: One 8.5x11 paper both sides can be used cheat sheet is allowed.

Problem 1 (4 Points)	3
Problem 2 (4 Points)	1
Problem 3 (4 Points)	2
Problem 4 (4 Points)	4
Problem 5 (4 Points)	4
Problem 6 (5 Points)	~~4~~ 5
============	
Total (25 Points)	19

(3)

Problem 1. (4 points)
Let $X_1, X_2, ..., X_n$ be independent random variables, each having a uniform distribution over (0,1). Let $M = \text{maximum}(X_1, X_2, ..., X_n)$.

- Show that the CDF of M, $F_M(x)$, is given by
$$F_M(x) = x^n, 0 \leq x \leq 1$$
- What is the probability density function of M, i.e. $f_M(m)$?

a) Each independent random variable $X_1 \cdots X_n$ has a uniform distribution over $(0,1)$. Denote the CDF of each X_i as x.

The CDF of M is defined as $P(\max(X_1, X_2 \cdots X_n) < r)$.
This is equivalent to $P[(X_1 < r) \cap (X_2 < r) \cdots (X_n < r)]$.

$$= P(X_1 < r) \cdot P(X_2 < r) \cdots P(X_n < r)$$

why?

$$= x_1 \cdot x_2 \cdots x_n$$

$$= x^n \quad \text{(because the distribution of all } X_i \text{ is equivalent)}$$

b) The PDF is the first derivative of the CDF. Hence,

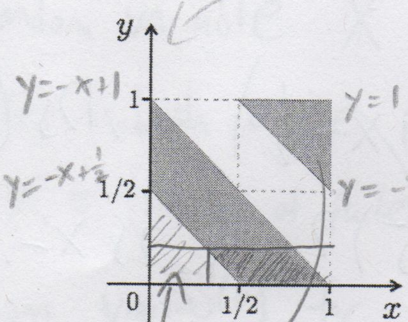
$$\text{PDF} = \frac{d}{dn}(x^n) = nx^{n-1}$$

1

Problem 2. (4 points)
A pair of jointly continuous random variables, X and Y , have a joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} c, & \text{in the shaded area shown in the figure below.} \\ 0, & \text{otherwise.} \end{cases}$$

$\frac{1}{4} = -x + \frac{1}{2}$



$0 \le X \le 1$

$\frac{1}{2} \le X \le 1$

Area of Shaded: $\frac{1}{2}$

(1) Find c.
(2) Find the marginal PDFs of X and Y, i.e., $f_X(x)$ and $f_Y(y)$.
(3) Find $E[X|Y = \frac{1}{4}]$ and $Var[X|Y = \frac{1}{4}]$, that is, the conditional mean and conditional variance of X given that $Y = \frac{1}{4}$.
(4) Find the conditional PDF for X given that Y = 3/4, i.e., $f_{X|Y}(x|Y = \frac{3}{4})$.

a) $\int_{x=0}^{1}\int_{y=-x+\frac{1}{2}}^{-x+1} c \, dy\, dx + \int_{x=\frac{1}{2}}^{1}\int_{y=-x+1.5}^{1} c \, dy\, dx$

$c\int \left(y\Big|_{-x+\frac{1}{2}}^{-x+1}\right) + c\int \left(y\Big|_{-x+1.5}^{1}\right)$

$c\int (\cancel{-x}+1+\cancel{x}-\frac{1}{2}) + c\int_{1/2}^{1} 1+x-1.5$

$c\int_0^1 \frac{1}{2} + c\int_{1/2}^{1} x-\frac{1}{2}$

$c\left(\frac{1}{2}x\Big|_0^{1}\right) + c\left(\frac{1}{2}x^2 - \frac{1}{2}x\Big|_{1/2}^{1}\right)$

$\frac{3}{8}c + \frac{1}{8}c = 1$ $\quad \boxed{c=2}$ ✓

see back →

b) $f_X(x) = \left(\int_{y=-x+\frac{1}{2}}^{-x+1} .375\, dy + \int_{y=-x+1.5}^{y=1} .125\, dy\right) 2$

$= .375y\Big|_{-x+\frac{1}{2}}^{-x+1} + .125y\Big|_{-x+1.5}^{1}$

$= .375\left(\cancel{-x}+1+\cancel{x}-\frac{1}{2}\right) + .125(1+x-1.5)$

$= (.125x - .25)2 = \boxed{.25x - .5}$

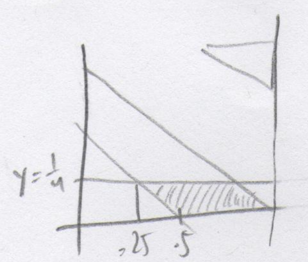
$y = -x + \frac{1}{2}$

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

$f_Y(y) = \left(\int_{x=\frac{1}{2}-y}^{1-y} .375\, dx + \int_{x=1.5-x}^{1} .125\, dx\right) 2$

$\left(.375x\Big|_{\frac{1}{2}-y}^{1-y} + .125x\Big|_{1.5-x}^{1}\right) 2$

c) Given the uniform joint PDF, the probability should be the area of the shaded parallelogram.



$\text{Area} = \left(\frac{1}{4} \times \frac{3}{4}\right) - 2\left(\frac{1}{4} \times \frac{1}{4} \times \frac{1}{2}\right) 2$

$\boxed{E[X|Y=\tfrac{1}{4}] = .25}$

2

Problem 3. (4 points)
The number of Packets arriving at a network router on a given day is Poisson distributed with mean $\lambda = 10$. The length of each packet to be a real-value uniformly distributed over $(0, 100)$ bytes. Find the mean and variance of the amount of data that the router receives in on a given day.

a) Suppose the random variable X denotes the total amount of bytes which arrive on a given day.
We know X = (number of packets) × (length of packets).
As the random variables for the number of packets and the length of packets are independent,

$E[X] = E[\text{num Packets}] \times E[\text{length of packets}]$

$E[\text{num Packets}] = \lambda = 10$

$E[\text{length of packets}] = \frac{0+100}{2} = 50$

$\boxed{E[X] = 10 \times 50 = 500}$ ✓

b) $Var[\text{num packets}] = \lambda = 10$

$Var[\text{length of packets}] = \cancel{\frac{(100-0)^2}{12}} = \frac{(100-0)^2}{12} = 833.33$

$\int_0^{100} \frac{1}{100} t \, dt = \frac{1}{2} \frac{1}{100} t^2 \Big|_0^{100} = \frac{100 \cdot 100}{2 \cdot 100} = 50$

$\int_0^{100} \left(\frac{1}{100}\right)^2 t \, dt = \frac{1}{2}\left(\frac{1}{100}\right)^2 t^2 = .5$

$Var(X) = E(X^2) - E(X)^2 \cancel{= .5 - 2500}$

$\boxed{Var[X] = 10 \times 833.33 = 8333.33}$ ✗

4

Problem 4. (4 points)

While running a computer simulation problem, the simulation is carried out in the form of N independent trials.

Each trial succeeds with probability p and fails with probability $(1-p)$ independent of previous trials. Your entire simulation ends upon the first successful trial.

Let the total time for the whole simulation, denoted by Y, be equal to the summation of time of all trials needed until the first success, and let X_i be the time needed for trial i where $\{X_i\}s$ are independent and identically distributed exponential random variables with parameter λ.

Using conditional transforms show that

$$Y^*(s) = G_N(X^*(s))$$

where

$$Y^*(s) \Leftrightarrow f_Y(y)$$

and

$$X^*(s) \Leftrightarrow f_X(x)$$

and

$$G_N(z) \Leftrightarrow p_N(n)$$

where $p_N(n)$ is the PDF for the number of trials N.

Y is a random sum defined as $\sum_{i=1}^{N} X_i$, where N is the geometrically-distributed number of trials until one succeeds and X_i is the exponentially distributed time of each trial. Suppose we take the Laplace transform of Y, conditioning on N.

$$Y^*(s \mid N=n) = E(e^{-sY} \mid N=n) = E\left(e^{-s\sum_{i=1}^{N} X_i}\right)$$

We can expand the term in the exponent and use the properties of e to further simplify.

$$E\left(e^{-s\sum_{i=1}^{N} X_i}\right) = E\left(e^{-s(X_1+X_2+\cdots+X_n)}\right) = E\left(e^{-sX_1} e^{-sX_2} \cdots e^{-sX_n}\right)$$

Since the terms within the brackets are independent

$$E\left(e^{-sX_1} e^{-sX_2} \cdots e^{-sX_n}\right) = E[e^{-sX_1}]\,E[e^{-sX_2}] \cdots E[e^{-sX_n}] = [X^*(s)]^n$$

next page

Finally, we need to uncondition on N.

$$Y^*(s) = \sum_{n=0}^{\infty} [X^*(s)]^n P[N=n] = G_n[X^*(s)]$$

Problem 5. (4 points)
Files stored on a computer can be either good or corrupted. In a good file the number of errors is Poisson distributed with mean 1; in a corrupted file it is Poisson distributed with mean 3. Suppose that any file's conditions depends on the condition of previous file only (assuming files are stored sequentially). Suppose that a good file is equally likely to be followed by either a good or a corrupted file, and that a bad file is twice as likely to be followed by a bad file as by a good file. Suppose that last file *call it file 0* was a good file.

(1) Find the expected total number of errors in the next two files (that is, in file 1 and 2).
(2) Find the long-run average number of errors per file.

a) $E(\text{errors} \mid \text{good file}) = 1$, $E(\text{errors} \mid \text{bad file}) = 3$

$P(\text{file}_{n+1} = \text{good} \mid \text{file}_n = \text{good}) = .5$

$P(\text{file}_{n+1} = \text{bad} \mid \text{file}_n = \text{good}) = .5$

$P(\text{file}_{n+1} = \text{good} \mid \text{file}_n = \text{bad}) = .33$

$P(\text{file}_{n+1} = \text{bad} \mid \text{file}_n = \text{bad}) = .66$

$P(\text{file 1} = \text{good}) = .5$, $P(\text{file 1} = \text{bad}) = .5$

$E(\text{errors in File 1}) = (.5 \times 1) + (.5 \times 3) = 2$

$P(\text{file 2} = \text{good}) = P(\text{file 1} = \text{good} \ \& \ \text{file 2} = \text{good}) + P(\text{file 1} = \text{bad} \ \& \ \text{file 2} = \text{good})$
$= (.5 \times .5) + (.5 \times .33) = .415$

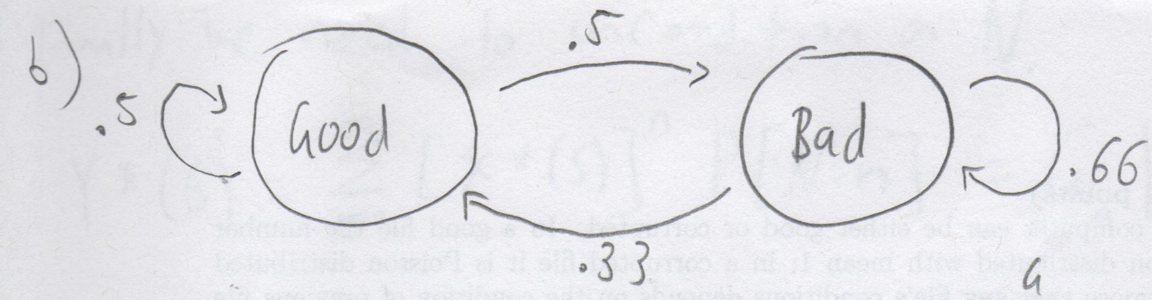
$P(\text{file 2} = \text{bad}) = P(\text{file 1} = \text{good} \ \& \ \text{file 2} = \text{bad}) + P(\text{file 1} = \text{bad} \ \& \ \text{file 2} = \text{bad})$
$= (.5 \times .5) + (.5 \times .66) = .58$

$E(\text{errors in File 2}) = (.415 \times 1) + (.58 \times 3) = 2.155$

$E(\text{errors in Files 1 and 2}) = E(\text{errors in File 1}) + E(\text{errors in File 2})$
$= 2 + 2.155 = \boxed{4.155}$ ✓

see back →

b)



$$\begin{array}{ccc} & G & B \\ \begin{array}{c} G \\ B \end{array} & \multicolumn{2}{c}{\begin{bmatrix} .5 & .5 \\ .33 & .66 \end{bmatrix}} \end{array}$$

$$.5G + .33B = G$$

$$.5G + .66B = B$$

$$.5G = .33B$$

$$G = \frac{.33}{.5} B$$

$$G + B = 1$$

$$\left(\frac{.33}{.5} + 1\right) B = 1$$

$$[G = .4 \quad B = .6]$$ Probabilities in the Long Run ✓

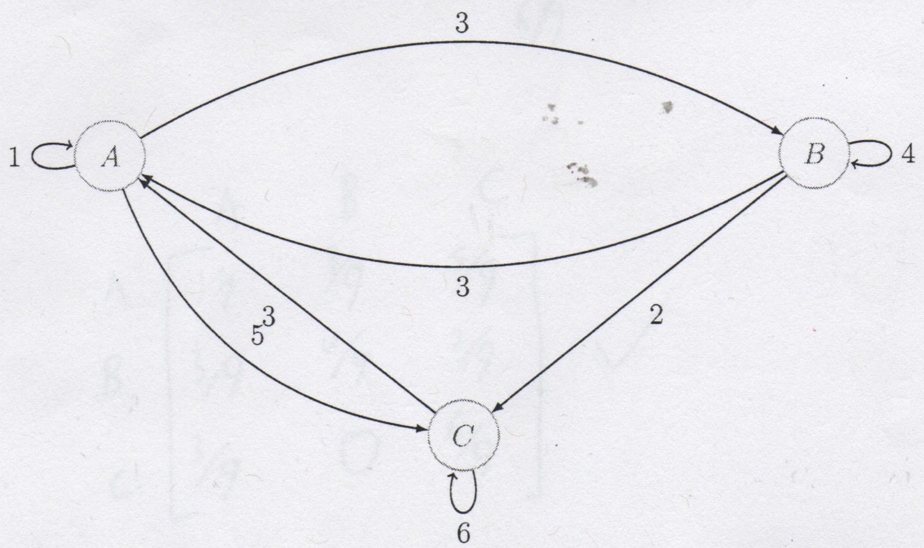
E(number errors per file in long run) $= (.4 \times 1) + (.6 \times 3) = \boxed{2.2}$

Problem 6. (5 points)
You are developing a search engine for your school website. You want to present the search results in the best way. Important pages should appear ahead of other less important pages in the search result. In order to find the rank"i.e, importance" of each page in the search result, you came up with this idea. Important pages are these pages that have more links from other pages linking to them.

The website has three page pages and the following graph represent their connectivity where the label the edge going from A to B indicates the number of links inside page A whose target is page B.

In order to define the importance of pages, you developed a web crawler that keeps traversing links between pages randomly. Each time, the crawler is at a given page and selects one of the outgoing links and follows it. The process is repeated indefinitely , you assume that the rank of a page is determined by the long term proportion of visits made to that page.

- Develop a DTMC that models the behavior of there states are the pages.
- draw the state transition diagram and list the state transition probability matrix P.
- Find the ranks of the three pages based on their connectivity shown in the graph below.



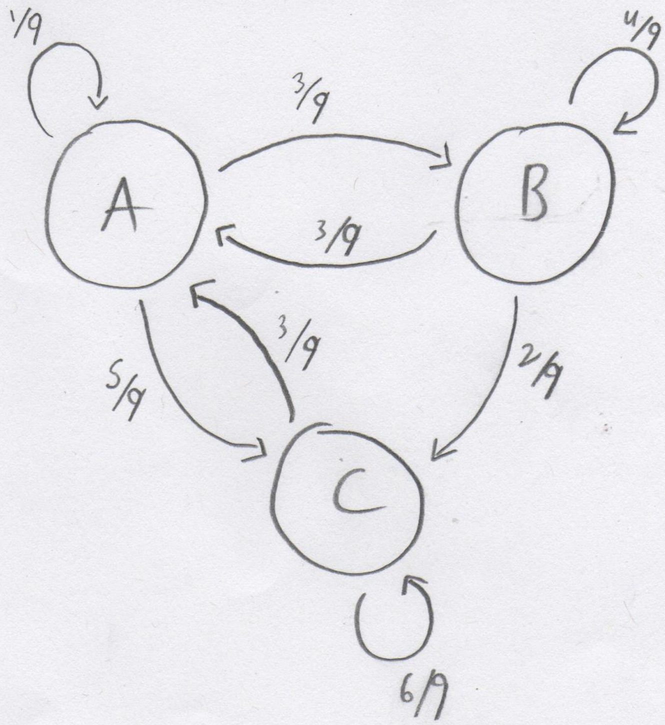
a) The states in this DTMC are the 3 pages. Since each page contains a certain number of links and one is chosen randomly when in the state, the probability of going from some state to another is the percent of links going to that site/state. This leads to the following table:

Percent Links to this Page

Starting page	A	B	C
A	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$
B	$\frac{3}{9}$	$\frac{4}{9}$	$\frac{2}{9}$
C	$\frac{3}{9}$	0	$\frac{6}{9}$

✓

b)



$$\begin{array}{c|ccc} & A & B & C \\ \hline A & 1/9 & 3/9 & 5/9 \\ B & 3/9 & 4/9 & 2/9 \\ C & 3/9 & 0 & 6/9 \end{array}$$

✓

c) $.11A + .33B + .33C = A$

$.33A + .44B = B$

$.55A + .22B + .66C = C$

$.33A = .56B$

$A = 1.7B$

$.935B + .22B + .66C = C$

$1.155B = .34C$

$C = 3.4B$

$A + B + C = 1$ ✓

$(1.7 + 1 + 3.4)B = 1$

$A = .28 \quad B = .16 \quad C = .56$

✓

Ranking

1) C

2) A

3) B