Final copy.

EE4-10
EE9CS5-1
EE9SC4

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2011

MSc and EEE PART IV: MEng and ACGI

PROBABILITY AND STOCHASTIC PROCESSES

Wednesday, 25 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL questions. All questions carry equal marks.

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : M.M. Draief

Second Marker(s) : R.B. Vinter

Probability and Stochastic Processes

1. We consider a random variable X taking positive integer values such that for $i \geq 1$,

$$\mathbf{P}(X = i) = \frac{2}{3^i}.$$

We define an integer-valued random variable Y as follows. Given $X = i$, the distribution of Y is uniform over the set $\{i, i+1\}$.

a) i) Check that the distribution of X defines a probability distribution.

Compute the mean value $\mathbf{E}(X)$ and the variance $\mathrm{var}(X)$ of X. [3]

ii) For $i \geq 1$, Compute $\mathbf{E}(Y \mid X = i)$. [3]

iii) Compute $\mathbf{E}(Y \mid X)$ and $\mathbf{E}(Y)$. [3]

b) i) Derive the joint distribution of the couple (X, Y). [3]

ii) Compute the marginal distribution of Y. [3]

iii) For $j \geq 1$, derive $\mathbf{E}(X \mid Y = j)$ and $\mathbf{E}(X \mid Y)$. [4]

iv) Compute $\mathrm{cov}(X, Y)$. [6]

2. Let $(X_n)_{n\geq 0}$ be a wide-sense stationary process with mean 0 and autocovariance function

$$c(m) = \text{cov}(X_n, X_{n+m}), \quad m \geq 0.$$

We are interested in the *best linear estimation* Y of X_{r+k}, $k > 0$, as a linear combination of $X_r, X_{r-1}, \cdots, X_{r-s}$, $s \geq 0$, i.e., $Y = \sum_{i=0}^{s} a_i X_{r-i}$ that minimises the mean squared error

$$f(a_1, \ldots, a_s) = \mathbf{E}\left[(Y - X_{r+k})^2\right].$$

a) Let $Y = \sum_{i=0}^{s} a_i X_{r-i}$ be the best linear estimator of X_{r+k}.

i) Show that

$$\mathbf{E}\left[(X_{r+k} - Y)X_{r-j}\right] = 0, \quad j = 0, \cdots s.$$

[4]

ii) Hence, prove that

$$\sum_{i=0}^{s} a_i c(|i-j|) = c(k+j), \quad 0 \leq j \leq s.$$

[6]

Hint: Distinguish the two cases $i \geq j$ and $j \geq i$.

b) Let $X_n = (-1)^n X_0$ where X_0 is equally likely to take each of the values -1 and 1.

i) Show that $(X_n)_{n\geq 0}$ is wide-sense stationary with zero mean and compute its autocovariance function. [2]

ii) Using 2.a), Show that the best linear approximation Y of X_{r+k} in terms of $X_r, X_{r-1}, \cdots, X_{r-s}$ is given by $(-1)^k X_r$. [2]

iii) Compute the mean squared error $\mathbf{E}\left[(Y - X_{r+k})^2\right]$. [2]

c) We now consider the following example, known as the *autoregressive process*, given by

$$X_n = \alpha X_{n-1} + Z_n, \quad n \in \mathbb{Z},$$

where $(Z_n)_{n\in\mathbb{Z}}$ are independent variables with 0 means and unit variances and where $|\alpha| < 1$. The autocovariance of $(X_n)_{n\in\mathbb{Z}}$ is given by

$$c(m) = \frac{\alpha^{|m|}}{1 - \alpha^2}, \quad m \in \mathbb{Z}.$$

i) Show that the best linear estimator of X_{r+k} is given by $Y = \alpha^k X_r$. [3]

ii) Show that the mean square error of the prediction is given by

$$\mathbf{E}\left((Y - X_{r+k})^2\right) = \frac{1 - \alpha^{2k}}{1 - \alpha^2}.$$

[6]

3. We study a model of the dynamics of the spread of an infectious disease in a population of healthy people of size N.

We assume that there are X_n infected individuals and $S_n = N - X_n$ healthy ones by day n. Between day n and day $n+1$, each of the S_n healthy individuals has a probability p of meeting a given infected individual and thus contracting the disease. Moreover, in day $n+1$ all the X_n infected individuals, in day n, recover and become healthy.

a) We assume that there are 3 individuals with $p = 1/3$.

i) Describe the transition matrix of the Markov chain X_n. [4]

ii) Classify the states of the chain according to whether they are recurrent or transient. [3]

iii) Does X_n has a stationary distribution? Comment. [4]

iv) Compute the average number of days before the epidemic stops starting with two infected individuals. [5]

b) We now assume that the population size is N and that the contact probability $p \in (0,1)$.

i) Justify the fact that if there are i infected individuals then the probability that a given healthy individual becomes infected in the following day is given by

$$1-(1-p)^i.$$

[3]

ii) Derive an expression for the transition probabilities

$$p_{ij} = \mathbf{P}(X_{n+1} = j \mid X_n = i).$$

[6]

4. We consider the following balls and bins problem.

We have M balls labeled $1, 2, 3, .., M$ and two bins A and B. At each time slot, we uniformly pick a random number $k \in \{1, 2, 3, \ldots, M\}$ and we remove ball k from the bin where it is and put it back in either A or B, chosen uniformly at random.

Let X_n denote the number of balls in bin A at time slot n.

a) Let $M = 4$.

i) Prove that $(X_n)_{n \geq 0}$ is a Markov chain. [2]

ii) Describe its transition matrix and draw its transition diagram. Justify your answer. [4]

iii) Is the chain irreducible? Is it aperiodic? [3]

iv) Derive its stationary distribution π. [3]

v) In this question, we assume that the number of balls is M, some positive integer, instead of 4.

Describe the content of bin A after running the dynamics described above for a long time. Justify your answer. [4]

b) In what follows, we suppose that $M = 4$.

i) We suppose that the bin A is initially empty, i.e. $X_0 = 0$. What is the probability that it contains an even number of balls after we run the above dynamics for a long time? Justify your answer. [4]

ii) We suppose that the bin A initially contains all four balls, i.e. $X_0 = 4$. We observe the evolution of the above Markov chain, i.e. $X_0, X_1, X_2, \ldots$. How often do we observe more balls in A than in B. [5]

SOLUTIONS - 2011 EE4-10 EE9 CS5-1 EE9 SC4 Q1

PROBABILITY & STOCHASTIC PROCESSES (2010-2011) 1/11

1/ X ; $\mathbb{P}(X=i) = \frac{2}{3^i}$ $i \geq 1$

Conditional on $\{X=i\}$; Y is uniform $\{i, i+1\}$.

a) i) $\boxed{\mathbb{E}(X) = 3/2 \; ; \quad \text{var}(X) = 3/4}$

Note: students can use the results of the course about geometric random variables

$$\mathbb{P}(X=i) = p\,(1-p)^{i-1} \quad ; \quad i \geq 1$$

$$\mathbb{E}(X) = \frac{1}{p} \quad \& \quad \text{Var}(X) = \frac{1-p}{p^2}.$$

Alternatively, they can compute them directly, ie.

$$\mathbb{E}(X) = \sum_{i\geq 1} i\, p(1-p)^{i-1} = p \sum_{i \geq 1} i\,(1-p)^{i-1}$$

$$= p \left[\frac{\partial}{\partial x} \sum_{i \geq 1} x^i\right]_{1-p}$$

$$= \frac{p}{(1-(1-p))^2} = \frac{1}{p}$$

& using similar technique for $\mathbb{E}(X^2)$.

ii) $i \geq 1$; $\mathbb{E}(Y \mid X=i) = \frac{1}{2} i + \frac{1}{2}(i+1) = \frac{2i+1}{2}$.

iii) Using previous question; $\mathbb{E}(Y \mid X=i) = \frac{2X+1}{2}$.

$$\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(Y \mid X)\big) = \frac{1}{2}\big(2\,\mathbb{E}(X)+1\big) = 2.$$

Q1 2

1/

b)

i) $i \geq 1$ $\quad p_{i,j} = \begin{cases} 0 & ; j \notin \{i, i+1\} \\ \frac{1}{3^i}\left(=\frac{1}{2}\frac{2}{3^i}\right) & ; j \in \{i, i+1\} \end{cases}$

ii) $p_j = \mathbb{P}(Y=j) = \begin{cases} 1/3 & ; j = 1 \\ \frac{4}{3^j} & ; j \geq 2 \end{cases}$

iii) First let us compute the conditional distribution X given $\{Y=j\}$.

- $Y=1$; $\quad X=1 \quad \Rightarrow \quad \mathbb{E}(X \mid Y=1) = 1$

- $Y=j>1$; $\quad X \in \{j-1, j\}$.

$$\mathbb{P}(X=j-1 \mid Y=j) = \frac{\mathbb{P}(X=j-1, Y=j)}{\mathbb{P}(Y=j)} = 3/4$$

Similarly; $\mathbb{P}(X=j \mid Y=j) = 1/4$

Hence, $\mathbb{E}(X \mid Y=j) = \frac{3(j-1)}{4} + \frac{j}{4} = \frac{4j-3}{4}$.

Finally; $\quad \mathbb{E}(X \mid Y) = \mathbb{1}_{\{Y=1\}} + \frac{4Y-3}{4}\mathbb{1}_{\{Y>1\}}$.

iv) $\quad \operatorname{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

$\mathbb{E}(X) = 3/2$; $\quad \mathbb{E}(Y=2)$

$$\mathbb{E}(XY) = \mathbb{E}\left(\mathbb{E}(XY \mid X)\right) = \mathbb{E}\left(X\,\mathbb{E}(Y \mid X)\right) = \mathbb{E}\left(X\,\frac{(2X+1)}{2}\right)$$

$$\Rightarrow \mathbb{E}(XY) = \frac{1}{2}\left[2\,\mathbb{E}(X^2) + \mathbb{E}(X)\right].$$

$\underline{Q\cancel{1}}$ 3

1/ b) iv) $\mathbb{E}(X^2) = var(X) + (\mathbb{E}(X))^2$

$$= \frac{3}{4} + \frac{9}{4} = 3.$$

$$\mathbb{E}(XY) = \frac{1}{2}\left(6 + 3/2\right) = \frac{15}{4}.$$

$$\boxed{cov(X,Y) = \frac{15}{4} - 3 = 3/4}$$

Q2

(4)

2/

$$f(a_1 .. a_s) = \mathbb{E}\left[(Y - X_{r+k})^2\right]$$
$$= \mathbb{E}\left(\left(\sum_{i=0}^{s} a_i X_{r-i} - X_{r+k}\right)^2\right)$$

a) i) $\frac{\partial}{\partial j} f(a_1 .. a_s) = 2 \mathbb{E}\left((Y - X_{r+k}) X_{r-j}\right) = 0$.

$j \in \{0, .., s\}$.

$\Rightarrow \mathbb{E}\left((Y - X_{r+k}) X_{r-j}\right) = 0$ as require!.

ii)

$$\mathbb{E}\left(\left(\sum_{i=0}^{s} a_i X_{r-i}\right) X_{r-j}\right) = \mathbb{E}(X_{r+k} X_{r-j})$$

$$(*) \sum_{i=0}^{s} a_i \mathbb{E}(X_{r-i} X_{r-j}). = c(k+j)$$

if $i \geq j$ $\quad r-j \geq r-j \quad \mathbb{E}(X_{r-i} X_{r-j}) = c(r-j-(r-i))$
$= c(i-j)$.

similarly $i \leq j$ $\quad \mathbb{E}(X_{r-i} X_{r-j}) = c(j-i)$

$\Rightarrow \mathbb{E}(X_{r-i} X_{r-j}) = c(|i-j|)$.

$$(*) \Rightarrow \sum_{i=0}^{s} a_i \, c(|i-j|) = c(k+j).$$

b) $X_n = (-1)^n X_0$.

i) $\mathbb{E}(X_n) = (-1)^n \mathbb{E}(X_0) = 0$.

$c(m) = ?$ $\quad \mathbb{E}(X_n X_{n+m}) = \mathbb{E}\left((-1)^{2n+m} X_0^2\right)$

since $\mathbb{E}(X_0^2) = 1/2 \, 1^2 + 1/2 \, (1)^2$, $\boxed{c(m) = (-1)^m}$
$= 1$

Q2

(5)

2(b)
ii)

By 2)(a) ii)

$$\sum_{i=0}^{s} a_i (-1)^{|i-j|} = (-1)^{k+j}$$

a solution is $a_0 = (-1)^j$; $a_i = 0$ $i \geq 1$

so that $Y = (-1)^k X_r$.

iii) mean squared error; note that $X_{r+k} = (-1)^{r+k} X_0 = (-1)^k X_r$

$$\Rightarrow \mathbb{E}\left(\left((-1)^k X_r - X_{r+k} \right)^2 \right) = 0$$

c) i) By 2)a)ii) $\sum_{i=0}^{s} a_i \alpha^{|i-j|} = \alpha^{k+j}$ $0 \leq j \leq i$

A solution is $a_0 = \alpha^k$; $a_i = 0$; $i \geq 1$.

$$\Rightarrow \boxed{Y = \alpha^k X_r}$$

ii) $$\mathbb{E}\left((Y - X_{r+k})^2 \right) = \mathbb{E}\left((\alpha^k X_r - X_{r+k})^2 \right).$$

$$= \mathbb{E}(X_{r+k}^2) - 2\alpha^k \mathbb{E}(X_r X_{r+k}) + \alpha^{2k} \mathbb{E}(X_r^2).$$

$$= c(0) - 2\alpha^k c(k) + \alpha^{2k} c(0)$$

$$= \frac{1}{1-\alpha^2} - \frac{2\alpha^{2k}}{1-\alpha^2} + \frac{\alpha^{2k}}{1-\alpha^2}$$

$$= \frac{1-\alpha^{2k}}{1-\alpha^2}$$

Q 3

6

3/ a) 3 individuals , $p = 1/3$.

Day n — Day $n+1$

X_n Infected — X_{n+1} depends only X_n

$N - X_n$ Heath — & not other past information.

i/ $(X_n)_{n \geq 0}$ is a Markov chain.

$X_n \in \{0, 1, 2, 3\}$

- $X_n = 0 \Rightarrow X_{n+1} = 0$

$$\boxed{\begin{array}{l} p_{00} = 1 . \\ p_{0j} = 0 \quad j = \{1, 2, 3\} \end{array}}$$

- $X_n = 1$; $X_{n+1} \in \{0, 1, 2\}$.

$p_{10} = \mathbb{P}(X_{n+1} = 0 \mid X_n = 1) = \mathbb{P}($no encounter, between infected & others$)$

$= 2/3 \cdot 2/3$

$$\boxed{p_{10} = 4/9}$$

$p_{12} = \mathbb{P}(X_{n+1} = 2 \mid X_n = 1) = \mathbb{P}($infected encounter, both healthy$)$

$= 1/3 \cdot 1/3 = 1/9$

$$\boxed{p_{12} = 1/9}$$

p_{11} $\mathbb{P}(X_{n+1} = 1 \mid X_n = 1) = 1 - (p_{10} + p_{12})$

$$\boxed{p_{11} = 4/9}$$

3/ a/ i/ $\boxed{7}$ Q 3

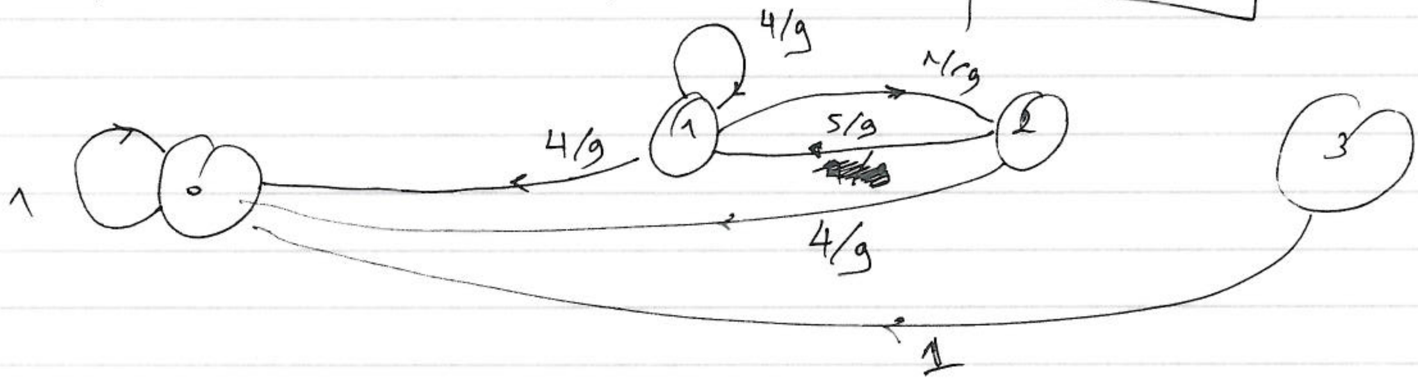
. $X_n = 2$, $X_{n+1} \in \{0, 1\}$.

$$p_{20} = \mathbb{P}(X_{n+1} = 0 \mid X_n = 2) = \mathbb{P}(\text{healthy does not encounter 2 infectious})$$

$$= 2/3 \cdot 2/3$$

$$\boxed{p_{20} = 4/9}$$

$$p_{21} = 1 - p_{20} = 5/9 \qquad \boxed{p_{21} = 5/9}.$$



$X_n = 3 \Rightarrow X_{n+1} = 0$

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 4/9 & 4/9 & 1/9 & 0 \\ 4/9 & 5/9 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

ii) $\{0\}$ recurrent

1, 2, 3 are transient

iii) The chain has a unique absorbing state 0 where the dynamics eventually ends (almost surely)

$$\Rightarrow \pi = (1, 0, 0, 0).$$

using $\pi P = \pi$ we also find that $\pi = (1, 0, 0, 0)$

Q3
$\frac{}{8}$

3/
a)

(iv) h_i = average number of days, starting from i before getting absorbed in 0

$h_0 = 0$; $h_1 = 1 + \frac{4}{9} h_0 + \frac{4}{9} h_1 + \frac{1}{9} h_2$.

$h_2 = 1 + \frac{4}{9} h_0 + \frac{5}{9} h_1$

$h_3 = h_0 = 1$.

$$h_2 = 1 + \frac{4}{9} + 5/9\, h_1 = \frac{13}{9} + 5/9\, h_1$$

$$\frac{5}{9} h_1 = 1 + \frac{4}{9} + \frac{1}{9}\left(\frac{13}{9} + 5/9\, h_1\right)$$

$$= 1 + \frac{4}{9} + \frac{13}{81} + \frac{5}{81} h_1.$$

$$\Rightarrow \quad \frac{40}{81} h_1 = \frac{81+36+13}{81} = \frac{130}{81} \Rightarrow \boxed{h_1 = \frac{13}{4}} = \frac{117}{36}$$

$$h_2 = \frac{13}{9} + 5/9\, \frac{13}{4} = \frac{52+65}{36} = \frac{117}{36}.$$

$$\boxed{h_1 = h_2 = \frac{13}{4}}$$

Q 3

9

3/b/i/

$X_n = i \Rightarrow X_{n+1} \in \{0, .., N-i\}$.

A health person at a given day will stay healthy in the following day if it does no encounter an infected individual which occurs with probability

$$(1-p)^i \quad \text{(by independence of meetings)}$$

Hence a healthy individual gets infected w.p.

$$1 - (1-p)^i$$

ii) By the previous argument, the number of heathy individuals who become infected is given by $\text{Bin}(N-i, 1-(1-p)^i)$

i being the number of infected individuals.

$$\Rightarrow \quad p_{ij} = \binom{N-i}{j}\left[1-(1-p)^i\right]^j (1-p)^{i(N-i-j)}.$$

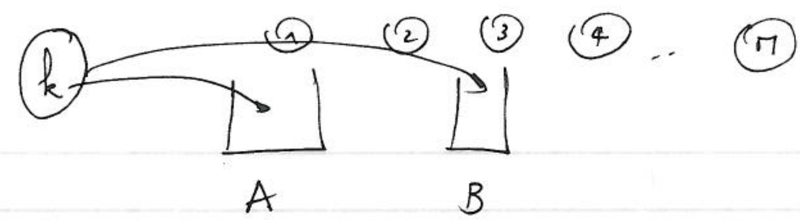
for $i+j \leq N$.

when $i+j > N$ $\quad p_{ij} = 0$.

Q 4

10

4/



a/

$M = 4$

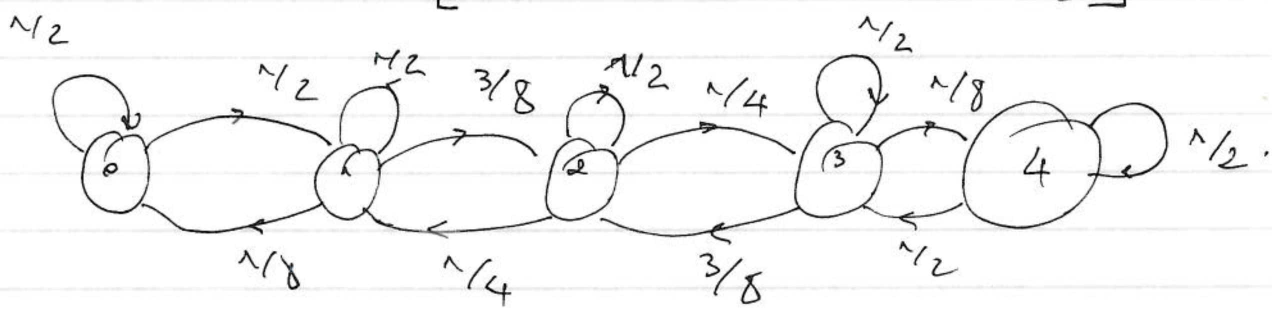
i/ The update of the content of the two urns is random, at each time step, yet it only depends on the contents of the urns in the previous time step only.

ii/

X_n: #balls in A

$X_n \in \{0, 1, 2, 3, 4\}$

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/8 & 1/2 & 3/8 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 3/8 & 1/2 & 1/8 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$



iii) As we can go from any state to any state with positive prob. then the chain is irreducible.

Moreover we can "loop" through the same state (as $p_{ii} \neq 0$ $\forall i$) so the chain is aperiodic

$\Rightarrow (X_n)_n$ is ergodic.

iv) By (iii) the chain has a unique stationary distrib. such that $\pi P = \pi \Rightarrow \cancel{\pi = \left(\frac{1}{16}, \frac{4}{16}, \frac{6}{16}\right.}, \frac{4}{16}, \frac{1}{16})$

$$\boxed{\pi = (1/16, 1/4, 1/2, 1/4, 1/16)}$$

Q4

11

iv) ~~In~~ general

the stationary distribution is determined by a binomial distribution $Bin(4, 1/2)$.

since $\pi_i = \binom{4}{i}\left(\frac{1}{2}\right)^i\left(1/2\right)^{4-i}$

v) As in previous question, one can anticipate that

$\pi \sim Bin(M, \frac{1}{2})$. (M general positive integer)

In fact, in the long run, each ball can be in either urn with equal probability 1/2 so that the content of Urn given by $X_n \underset{n\to\infty}{\longrightarrow} X_\infty \sim Bin(\pi, 1/2)$.

b) BACK TO $\pi = 4$:

i) $X_0 = 0$ $\mathbb{P}$(even number of balls in the long run)

$= \lim_n \mathbb{P}(X_n \in \{0, 2, 4\}) = \pi_0 + \pi_2 + \pi_4$
$= 1/2$.

ii) $X_0 = 4$

$\mathbb{P}(X_n \in \{0, 1\}) \underset{n\to\infty}{\longrightarrow} \pi_0 + \pi_1 = \frac{5}{16}$.