UNIVERSITY OF LONDON

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2003

This paper is also taken for the relevant examination for the Associateship.

M3S4/M4S4 APPLIED PROBABILITY

DATE: Tuesday, 3rd June 2003 TIME: 2 pm – 4 pm

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used. Statistical tables will not be available.

1. *a)* If $F(x)$ is the distribution function of X, a continuous random variable which takes only positive values and for which $E(X)$ exists, prove that

$$E(X) = \int_0^\infty [1 - F(x)]\,\mathrm{d}x\ .$$

b) In a simple Poisson process, the number of events which occur in a time interval of length t is $Poisson\,(\lambda t)$.

i) From this, derive the probability density function (pdf) of the waiting time, T, between consecutive events.

ii) State the mean of T.

iii) The kth event will occur in the interval $[t, t + \delta t]$ if and only if $k - 1$ events occur in the interval $[0, t]$ and one event occurs in the interval $[t, t + \delta t]$. Use this fact to derive the probability density function of the time to the kth event.
What is the name of the associated distribution?

c) The rate at which items are purchased at a small shop is modelled as a compound Poisson process. Customers arrive at the checkout according to a Poisson process with a mean μ arrivals per hour, where they are immediately served. The number, X, of purchases made by each customer is modelled as a geometric random variable, having

$$P(X = x) = q^{x-1}p, \quad x = 1, 2, \ldots; \quad p + q = 1.$$

Derive the probability generating function of Z, the number of items purchased per hour, and hence find the mean number of items purchased per hour. You may quote, without derivation, the probability generating functions of the Poisson and geometric distributions, and you may use, without derivation, an appropriate way of combining these.

2. *a)* *i)* If a Markov chain is said to be *irreducible*, what does this mean?

ii) Write down and explain the meaning of the Chapman-Kolmogorov equations.

iii) Define the terms *transient*, *recurrent*, *null recurrent*, and *positive recurrent* for Markov chains.

b) In a psychological study, an individual is classified as being in one of three possible moods: euphoric (state 0), normal (state 1), or depressed (state 2). Observation of the individual over an extended period of time shows that the transition matrix for his probabilities of moving between states from one day to the next is

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

i) If he is in state 0 on Monday, find the probability that he will be in state 0 on Wednesday.

ii) Find the probability that, in the long term, he will be in each of states 0, 1, and 2.

iii) Find the mean recurrence time of state 0.

c) The *Ehrenfest chain* originated in physics, but can be modelled as a system of N balls distributed between two urns, A and B. We pick one of the N balls at random and move it to the other urn.

i) Write down the transition matrix for the number of balls in urn A when $N = 3$.

ii) Draw the transition diagram of this matrix.

iii) Why does this system not have a limiting distribution?

3. *a)* At each generation, individual organisms in a population act independently to split into X offspring, where X has a $Poisson\,(\mu)$ distribution.
The population starts from a single organism at generation 0.

i) Find the mean size of the nth generation.

ii) Find the expected total number of organisms that have existed altogether up to and including the nth generation.

iii) If observation suggests that $P(X = 0) \approx 1/\sqrt{e}$, what would you expect the mean generation size to be in the long term?

iv) If it is known that $\mu = 2\ln 2$, show that the probability of ultimate extinction of this process is 0.5.

v) If, instead of starting with a single individual, the process in (*iv*) starts with M individuals, how large must M be to ensure that the probability of extinction is less than 1/32?

b) Suppose that, in a branching process with immigration, the mean of the offspring size probability distribution is μ and the mean number of immigrants per generation is ν. Find an expression for μ_n^*, the mean number of individuals in the nth generation, in terms of μ_{n-1}^*.
[You may use without derivation results relating to probability generating functions.]

4. *a)* A particular stochastic process has led to the following partial differential equation for the probability generating function $\Pi(s,t)$,

$$t\frac{\partial \Pi}{\partial s} = s\frac{\partial \Pi}{\partial t} + st\frac{\partial}{\partial s}(s\Pi) \ .$$

Solve this equation for $\Pi(s,t)$ using the initial condition that $\Pi(s,0) = s^2$.

b) *i)* In a general birth and death system which currently contains x objects, the overall birth rate of new objects is β_x and the overall death rate is ν_x. Write down or derive the differential difference equations for $p_x(t)$, the probability that there are x objects in the system at time t.

ii) Cells in a particular biological system have an exponential lifetime distribution with parameter ν. The larger the system is, the less likely it is to accumulate new cells, and in fact new cells are added to the system according to a Poisson process with parameter $\beta/(1+x)$ where x is the number of cells already in the system. Use the general differential difference equation from part (i) to give differential difference equations for $p_x(t)$, the probability that there are x cells in the system at time t. Briefly describe, without actually performing the calculations, how you could derive a partial differential equation for the probability generating function of the number of cells at time t from these differential difference equations.

5. Two candidates, A and B, stand for election. In the population of voters, a proportion p will vote for A, a proportion q for B, and a proportion r will abstain, with $p+q+r=1$. Let $n_A(t)$ be the number who have voted for A by time t and $n_B(t)$ be the number who have voted for B by time t, where the votes are recorded sequentially at random. The winner is declared when a candidate first achieves M more votes than the rival candidate.

An attempt to rig the election was made by recording bogus votes for the candidates before the voting started. Suppose that A has j more bogus votes than B (where j may be negative), so that, by time t, the number of votes apparently favouring A is $m(t) = j + n_A(t) - n_B(t)$.

a) Let q_j be A's probability of losing the election. You can assume that the population is so large that it can be regarded as effectively infinite. By considering the first legitimate vote (i.e. after time $t=0$), which could be for A or B or an abstention, derive the relationship

$$q_j = q_{j+1}p + q_j r + q_{j-1}q \ .$$

b) Solve this recurrence relation for $p \neq q$ and for $p = q$.

[You may wish to use the result that the solution of the recurrence relation $x_j = ax_{j+1} + bx_{j-1}$, with $0 < a, b < 1$, is

$$x_j = \begin{cases} c_1 + c_2(b/a)^j & a \neq b; \\ c_3 + c_4 j & a = b, \end{cases}$$

where the c_i are constants determined by boundary conditions.]

c) If D_j is the expected number of votes which are cast before an outcome is reached, show that

$$(p+q)\,D_j = pD_{j+1} + qD_{j-1} + 1 \ .$$

d) Hence, for the case when $p = q$, show that a solution is given by

$$D_j = c_5 + c_6 j - j^2/2p.$$

State the boundary conditions, and hence find the constants c_5 and c_6.