UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2004

BEng Honours Degree in Computing Part III
BEng Honours Degree in Information Systems Engineering Part III
MEng Honours Degree in Information Systems Engineering Part III
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the
Associateship of the City and Guilds of London Institute
This paper is also taken for the relevant examinations for the
Associateship of the Royal College of Science

PAPER C337=I3.18

SIMULATION AND MODELLING

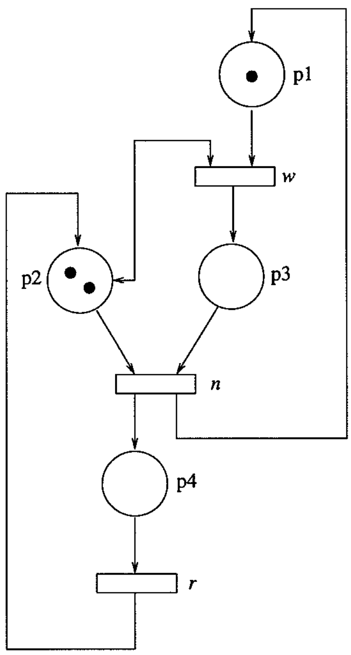
Friday 7 May 2004, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 The Imperial Country Club offers golf lessons with its two resident professionals ("pros"). The lessons are very popular and there is always a willing client in the clubhouse ready to sign up for a lesson. Before the lesson starts, the client spends a short time hitting balls into a practice net which is specially fitted with a video camera and monitor. This enables the pro to point out basic faults which will be addressed during the main part of the lesson. There is only one practice net, so if the second pro is available whilst the other is with a client in the net, he has to wait until the net session is over before the next lesson can begin. At this point the club receptionist calls for the next waiting client by making an announcement over the clubhouse tannoy system. On average t_w minutes pass between the client being called and their completing the walk from the clubhouse to the practice net. After a net session the pro and client move out to the outdoor practice-range where the main lesson begins. The practice range is large enough for both pros to be giving lessons at the same time should the need arise. A net session lasts on average t_n minutes and a practice range session on average t_r minutes.

The system can be represented abstractly in the form of a Petri net as shown in the figure below. A transition labelled 'x' has associated delay $t_x, x \in \{w, n, r\}$.



a In relation to the problem specification, what interpretation can be placed on the markings of the four places p1, p2, p3, p4? What is the purpose of the double-headed arrow between w and p2?

b Draw the reachability graph for this Petri Net.

c Assume that each of the three time delays is exponentially distributed and let $r_x = 1/t_x$ denote the firing rate associated with transition x, $x \in \{w, n, r\}$. In this case the underlying transition system corresponds to a Markov chain. From the reachability graph, or directly from the original problem description, draw the Markov process for the system, labelling the transitions r_w, r_n or r_r accordingly.

d From the Markov process, or otherwise, derive the generator matrix $\mathbf{Q}$.

e The club has decided to invest in a second practice net in order to allow both pros to administer net sessions simultaneously. Draw a modified Petri net that permits this behaviour, including any changes required to the initial marking. You are *not* required to produce the reachability graph or underlying Markov process.

(The five parts carry, respectively, 25%, 25%, 15%, 15% and 20% of the marks).

2 A print service operated by a computing department uses four printers. Two of the printers are dedicated to processing large jobs (> 30 pages) and the other two for small jobs (≤ 30 pages). The number of pages in each print job has been measured over a period of time and the print size distribution has been captured in the form of a distribution sampler `psize` such that `psize.next()` will return a sample job size, measured in pages. Whether each page is single- or double-sided is not relevant - it is the number of pages that matters.

For each class of job (i.e. large or small), there are thus two choices of printer. The print server, which allocates jobs to printers, always selects the printer that has the smallest *number* of outstanding print jobs. Once the job has been allocated to a printer it joins (the back of) a first-come-first-served (FCFS) queue associated with that printer. The queues have very large capacity to the extent that they can be assumed never to be full. Each printer can print one page every two seconds.

The inter-arrival times of the print jobs from the users has also been measured and captured in the form of a distribution sampler `iat`. The call `iat.next()` returns a sample from this distribution.

a If the print-job sizes and inter-arrival times are both exponentially distributed, is it possible to analyse this system using simple queueing theory? Explain your answer.

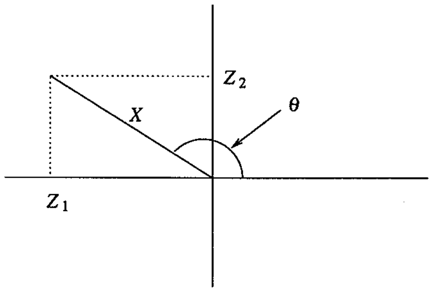
b Design a discrete-event simulation of this system which models the arriving print jobs and printers and which measures the mean and variance of the response times (i.e. queueing time + print time) for both large and small jobs. To compute the latter, use a `Measure` object with methods `void add( double x )` to add an observation, and methods `double mean()` and `double variance()` to return the sample mean and variance. You do not need to model the print server explicitly as it operates many orders of magnitude faster than the printers.

You may use any Java-like notation you wish and may assume either event scheduling or process interaction models of discrete-event simulation. Do not detail obvious supporting library methods/functions. Avoid excessive detail.

c Now assume that the choice of which queue to join is governed by the time-to-completion of the print jobs, rather than the total number of jobs. The printer which minimises this will be chosen by the print server. Explain in general terms how you would modify your code to accommodate the new assumption. You may wish to sketch some code but you do not need to detail a complete implementation.

(The three parts carry, respectively, 10%, 70% and 20% and of the marks).

3a In the diagram below Z_1 and Z_2 are independent random variables normally distributed with mean 0 and standard deviation 1. It can be shown that $X^2 (= Z_1^2 + Z_2^2)$ in the diagram is exponentially distributed with parameter $\frac{1}{2}$ and that the angle θ is uniformly distributed on the interval $(0,2\pi)$. Using this information develop a method for sampling a normal distribution with mean μ and standard deviation σ given a method for sampling an exponential distribution with parameter λ. Credit will be awarded if your method is 1:1 i.e. if it produces one (normal) sample for each random number generated. You may wish to sketch some code for implementing the method in order to demonstrate that the 1:1 property holds.



b If $Y \sim N(\mu, \sigma^2)$ then $X = e^Y$ has a *lognormal* distribution with a mean of

$$\mu_X = e^{\mu+\sigma^2/2}$$

and variance

$$\sigma_X^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

i. Show that μ and σ^2 can be expressed in terms of μ_X and σ_X^2 thus:

$$\mu = \log \frac{\mu_X^2}{\sqrt{\mu_X^2 + \sigma_X^2}} \qquad \sigma^2 = \log\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)$$

Hint: Derive the equation for σ_X^2 first. Note that σ_X^2 can be written $\mu_X^2(e^{\sigma^2} - 1)$ and recall that $e^{\log x} = x$.

ii. Using this, show how an existing generator for sampling a normal distribution (e.g. the one from part a) can be used to sample a lognormal distribution with a specified mean m and variance v.

c With the aid of a diagram explain how the rejection method can be used to sample a distribution with density function $f(x)$, where $a \le x \le b$ and where

$$\max_{a \le x \le b} f(x) = m$$

As part of your answer explain, with either formal or informal arguments, why the method produces samples with the required distribution. On average how many $U(0,1)$ random numbers will be needed to generate one sample from the distribution? Hint: if a random variable X has probability distribution function $P(X = x) = q^x(1-q)$ then $E[X] = q/(1-q)$.

(The three parts carry, respectively, 40%, 30% (20+10) and 30% of the marks).

4a What is meant by an *unbiased* estimate of an unknown quantity α? If $X_i, 1 \leq i \leq n$ are observations each with finite mean α, show that the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is an unbiased estimate of α.

b A simulation is used to produce n estimates, $X_i, 1 \leq i \leq n$, each with unknown mean α. Let the sample mean and variance of the estimates be $\overline{X}$ and S^2 respectively. Explain what is meant by a *95% confidence interval* for α and show how $\overline{X}$ and S^2 can be used in conjunction with tables to compute such a confidence interval. Why is it important that the X_i are independent and normally distributed for the confidence interval to be exact?

c Describe the *batched means* method of generating n observations $X_1, ..., X_n$, from a single run of a simulation. Explain why the X_i, might not be independent in some cases and why, if independence is assumed when the X_i are *not* independent, the computed confidence intervals will typically be narrower than they should.

d In a particular simulation, confidence intervals for the estimated mean-time-to-failure of a manufacturing system have been gathered for different values of the load parameter L (jobs completed per hour). For each run the simulation was executed for the equivalent of ten simulated days. Eight independent replications, each with the same value of L, were used to generate the confidence interval for that value of L.

When analysing the results, the width of the confidence interval is seen to be substantially wider on runs with high values of L when compared to those with low values of L.

i) Give one explanation of why the confidence intervals might be behaving the way they are. If each of the eight simulations (for a given value of L) is executed for a longer time, e.g. twenty simulated days, why will this reduce the width of the confidence interval?

ii) Suppose instead that the simulation is executed more than eight times for each value of L but with the same run length as before (ten simulated days). Explain why this too will reduce the width of the confidence interval. How many independent executions would be needed to halve the width of the confidence interval on average? Explain your answer.

(The four parts carry, respectively, 20%, 25%, 25% and 30% (15+15) of the marks).