UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1999

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
Diploma of Membership of Imperial College
Associateship of the Royal College of Science
Associateship of the City and Guilds of London Institute

PAPER 3.37 / I 3.18

SIMULATION AND MODELLING

Thursday, April 29th 1999, 10.00 – 12.00

Answer THREE questions

For admin. only:
paper contains 4 questions

1 A complex telecommunication network is being designed to route packets from a set of network inputs to a set of network outputs. It works by routing the packets through a series of intermediate switching nodes although the exact manner in which this is done is unimportant.

A simulation model of the network has been constructed with a view to optimising the placement of the intermediate switches, each of which consists of a single FIFO queue at each input and a routing matrix to feed packets from the input queues to the switch outputs. Again the exact manner in which this is done is unimportant.

Although the simulation accurately models the network behaviour and input processes, it contains no measurement code at all (not even in the various libraries used to build the simulation). You have been asked to advise on how best to incorporate such code.

In the context of this example, explain how you would go about measuring the following:

a The mean number of packets at a given switch input queue waiting to be forwarded

b The mean and variance of the number of intermediate nodes that a packet takes en route from source to destination

c The distribution of the packet transmission time (the transmission time of a packet is the total time the packet spends in the network).

d The probability that the transmission time is greater than some given value t.

e The distribution of the total number of packets in the network at any time

As part of your answer you should outline any measurement variables, data structures, events etc. that your methods require. Any assumptions you need to make about the operation of the network or the simulation should be clearly stated. The efficiency of the methods you describe will be taken into consideration when awarding credit.

Note: The five parts carry equal marks.

2 An analysis of the operation of a manufacturing company has revealed that the down time of its machines has approximately the Gamma(2,θ) distribution where θ depends on the machine; samples from this distribution are required in order to implement a simulation model of the business. The cumulative distribution function and density function of the Gamma(2,θ) distribution are respectively (for $x \geq 0$):

$$\begin{aligned} F(x) &= 1 - e^{-2\theta x} - 2\theta x e^{-2\theta x} \\ f(x) &= 4\theta^2 x e^{-2\theta x} \end{aligned}$$

a Describe how the *transform* method can be used to generate samples from a distribution stating any mathematical properties which the method assumes. As part of your answer explain why the method cannot be used to sample the Gamma(2,θ) distribution.

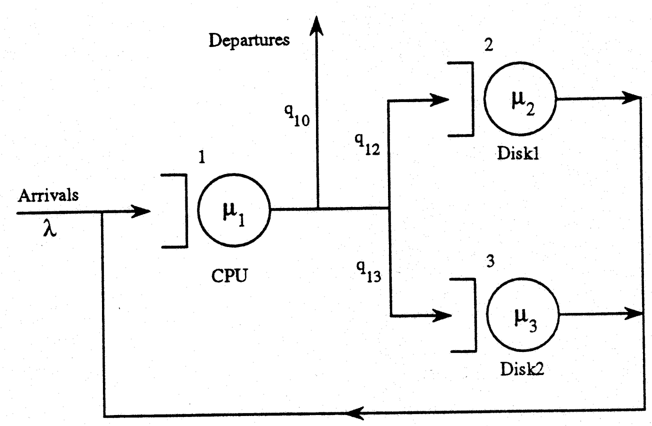
b For a given machine, θ has been estimated to be 1. For this parameterisation describe how the rejection method can be used to sample the Gamma(2,θ) distribution. As part of your answer, compute the height, m, of the bounding box of the density function and explain how you would determine the value, v, on the x-axis such that 99% of the distribution will be sampled.

c For part b write down an expression in terms of m and v for the probability that a sample will be generated on the first attempt. What will be the expected number of rejections before each sample is generated?

Note: The three parts carry 30%, 40% and 30% of the marks respectively.

3 A CPU executes a series of jobs each of which requires occasional accesses to disk. Disk requests are buffered in a FCFS manner at the disk controllers; the disk buffers are very large and are well approximated by infinite-capacity queues. When an IO request is sent to a disk, the CPU switches to the next job in the job queue (if there is one). When a job's disk IO is complete, the job is re-enabled for execution by re-entering it into the job queue. When a job is complete it leaves the system.

The system in question has two disks and can be represented by a queueing network as shown below:



From observations the average workload has been established as follows:

- Each submitted job makes 70 visits to disk 1 and 50 to disk 2 *on average*
- The mean service times are 0.005s for the CPU, 0.03s for disk 1 and 0.027s for disk 2
- The arrival rate of new jobs to the system is $\lambda = 0.4$

Assuming that the inter-arrival times of new jobs to the system, and the service times at the CPU and disks are approximately exponentially distributed,

a What are the routing probabilities q_{10}, q_{12}, q_{13} shown on the diagram? (Note that $q_{10} + q_{12} + q_{13} = 1$.)

b By solving the corresponding traffic equations, or otherwise, compute the throughput and utilisation of the CPU and of each disk

c Determine the mean *total* number of jobs in the system. Hence, using Little's Law or otherwise, determine the mean processing time for a job (the mean total time spent in the system).

d The owner of the system is contemplating replacing the CPU with one five times faster. By what factor (stated as a percentage decrease) will the mean processing time be reduced if this is done? Give a short exaplanation of your findings.

Note: The four parts carry 10%, 40%, 30% and 20% of the marks respectively.

4a It is required to model a client-server database system and as part of the exercise the times between successive client requests have been measured over an extensive period of time. When simulating the system either the measured data can be used as input, or the data gathered can be fitted to a known mathematical distribution and the distribution sampled instead. Give *two* reasons why distribution sampling is preferable in this type of situation.

4b For a particular set of measured data it is suspected that the data are samples from a $U(0,b)$ distribution, but the value of b is not known and so must be estimated. The maximum likelihood estimator for b in this case can be shown to be $\hat{b} = \frac{n+1}{n}M$ where $M = max_{1 \leq i \leq n} X_i$ and X_i is the i^{th} data sample, $1 \leq i \leq n$. The measured data are as follows:

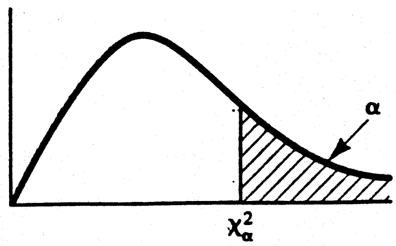
1.25	5.49	7.70	2.77	5.35	5.17	0.90	4.85	6.99	4.26
4.66	3.61	5.74	7.66	6.11	2.99	4.95	0.82	2.37	4.83
1.65	4.44	5.75	0.56	1.61	2.60	3.10	1.23	5.47	3.75
7.29	7.51	0.07	6.91	2.81	0.69	4.84	4.84	3.50	7.47
7.32	6.72	0.40	3.92	5.90					

i Construct a five-bucket histogram from this data and compute the expected contents of each of the five buckets.

ii Perform a χ^2 test at the 2.5% significance level in order to test the null hypothesis, H_0 : the observations are $U(0,b)$ distributed. Show your working.

Note: The two parts (a and b) carry 30% and 70% of the marks respectively.

PERCENTAGE POINTS OF THE CHI-SQUARE DISTRIBUTION WITH V DEGREES OF FREEDOM



ν	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$t^2_{0.10}$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78
5	16.7	15.1	12.8	11.1	9.2
6	18.5	16.8	14.4	12.6	10.6
7	20.3	18.5	16.0	14.1	12.0
8	22.0	20.1	17.5	15.5	13.4
9	23.6	21.7	19.0	16.9	14.7
10	25.2	23.2	20.5	18.3	16.0
11	26.8	24.7	21.9	19.7	17.3
12	28.3	26.2	23.3	21.0	18.5
13	29.8	27.7	24.7	22.4	19.8
14	31.3	29.1	26.1	23.7	21.1
15	32.8	30.6	27.5	25.0	22.3
16	34.3	32.0	28.8	26.3	23.5
17	35.7	33.4	30.2	27.6	24.8
18	37.2	34.8	31.5	28.9	26.0
19	38.6	36.2	32.9	30.1	27.2
20	40.0	37.6	34.2	31.4	28.4
21	41.4	38.9	35.5	32.7	29.6
22	42.8	40.3	36.8	33.9	30.8
23	44.2	41.6	38.1	35.2	32.0
24	45.6	43.0	39.4	36.4	33.2
25	49.6	44.3	40.6	37.7	34.4
26	48.3	45.6	41.9	38.9	35.6
27	49.6	47.0	43.2	40.1	36.7
28	51.0	48.3	44.5	41.3	37.9
29	52.3	49.6	45.7	42.6	39.1
30	53.7	50.9	47.0	43.8	40.3
40	66.8	63.7	59.3	55.8	51.8
50	79.5	76.2	71.4	67.5	63.2
60	92.0	88.4	83.3	79.1	74.4
70	104.2	100.4	95.0	90.5	85.5
80	116.3	112.3	106.6	101.9	96.6
90	128.3	124.1	118.1	113.1	107.6
100	140.2	135.8	129.6	124.3	118.5