Computer Systems Modelling

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

Computer Science Tripos, Part II

Michaelmas Term 2004

R. J. Gibbens

Problem sheet

William Gates Building
JJ Thomson Avenue
Cambridge
CB3 0FD

http://www.cl.cam.ac.uk/

Simulation

1. In what contexts is simulation an appropriate technique for performance evaluation? When it is inappropriate?

2. How would you generate random variates for the exponential distribution? Write pseudo code for a simulation component which models the arrival process of customers at a bank, given that the mean arrival rate is 40 customers per hour. What are the important events for this component (the generator).

3. How would you generate random variates from the mixed exponential distribution

$$f(x) = 1/3\lambda_1 \mathrm{e}^{-\lambda_1 x} + 2/3\lambda_2 \mathrm{e}^{-\lambda_2 x}$$

   where $\lambda_1 = 5, \lambda_2 = 10$, which describes the joint arrival process of two separate exponentially distributed streams of customers to a single queue?

4. Suppose that a simulation is constructed to estimate the mean response time, in milliseconds, of an interactive computer system. A number of repetitions are performed with the following results (measured in milliseconds): 4.1, 3.6, 3.1, 4.5, 3.8, 2.9, 3.4, 3.3, 2.8, 4.5, 4.9, 5.3, 1.9, 3.7, 3.2, 4.1, 5.1.

   Calculate the sample mean and variance of these results and hence derive a 95% confidence interval for the mean response time. You may wish to make use of the fact that $\mathbb{P}(Z > 1.96) > 0.025 > \mathbb{P}(Z > 1.97)$ where Z is a random variable with the unit normal distribution.

5. An experiment is performed to estimate the performance of a $M/M/1$ system with FIFO queueing. The response times of each of 1000 successive customers are recorded and the sample mean ($\overline{X}$) and sample variance (S^2) of these numbers are calculated.

   A naïve student believes that $100(1-\alpha)\%$ confidence bounds for the mean can be derived by the expression

$$\overline{X} \pm \frac{z_{\alpha/2} S}{\sqrt{1000}}$$

   where $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$ and Z is a standard Normal random variable.

   What mistake has this student made? Would their technique be valid if they were estimating the service time of this queue?

6. Using the inverse transform method show that

$$X = \left\lfloor \frac{\log(U)}{\log(1-p)} \right\rfloor + 1$$

   has a geometric distribution with parameter p where U has the $U(0, 1)$ distribution.

7. Describe a procedure for the generation of the first T time units of a Poisson process of fixed rate λ.

8. Describe the variance reduction technique based on antithetic variables and give an example of how it is used.

9. For the variance reduction technique based on control variates derive the optimal choice of $c = c^*$ and its associated variance given in the lectures.

Operational analysis

1. In what kinds of context are the tools of operational analysis appropriate for performance evaluation? When are they inappropriate?

2. A printer and print queue are observed for 48 minutes during which time 16 jobs were completed. The mean queue length was 4 jobs. The printer could print 8 pages per minute. The mean job size was 12 pages.

   Assuming that the number of arrivals was equal to the number of completions, calculate

   (a) the throughput of the system in jobs-per-minute,

   (b) the mean residence time of a job,

   (c) the mean queuing time of a job, and

   (d) the utilization of the printer.

3. An queueing network described as *open* and *feed forward* is to be analyzed using operational laws. The mean arrival rate into the system is known to be λ.

   (a) Describe the significance of the two phrases in italics.

   (b) The mean service times, visit counts and connectivity are known for each device. Why is it still impossible, in general, to calculate an upper bound on the residence time of a job in the network?

   (c) Suppose, however, that there are never more than five jobs in the network at once. Outline how you could derive an upper bound on the residence time.

4. Suppose that busses run according to a regular timetable with a fixed inter-arrival time of 10 minutes. What is the mean waiting time of a customer arriving at the bus stop?

   Suppose instead that bus inter-arrival times are exponentially distributed, still with mean 10 minutes. What now is the mean waiting time of an arriving customer?

   Derive a general relationship between the mean waiting time and the first and second moments of the inter-arrival distribution. *Hint: if the inter-arrival time pdf is $f(x)$ then note that the probability of encountering an interval of length τ is proportional to both $f(\tau)$ and to τ.*

Queueing theory

1. In what situations does queuing theory provide appropriate techniques for performance evaluation? When does it not?

2. Show for the M/M/1 queue that the probability that there are n or more customers in the system is given by ρ^n.

   Use this result to find a service rate μ such that, for given λ, n, α where $0 < \alpha < 1$, the probability of n or more customers in the system is given by α.

   Find a value of μ for an M/M/1 queue for which the arrival rate is 10 customers per second, and subject to the requirement that the probability of 3 or more customers in the system is 0.05

3. What is the average length of each idle period of an M/M/1 server, given its arrival rate λ and service rate μ. What is the average length of a busy period?

4. Using the steady-state distribution of the number of jobs in a $M/M/1$ queueing system given in lectures derive the first and second moments of this distribution and hence the variance of the number of jobs present. Describe what happens as the load ρ increases.

5. Given an $M/M/1/K$ queue with $\lambda = 10$, $\mu = 12$ and $K = 15$ over what proportion of time are customers rejected from the queue? What is the effective arrival rate? What is the effective utilization of the server?

6. Repairing a computer takes 4 stages in sequence, namely removing the lid, finding the faulty part, replacing it, and reassembling the machine. Each step is independent and exponentially distributed with mean 3 minutes. What is the coefficient of variation of the repair time? Construct a Markov chain model of this system, assuming an infinite population of machines.

7. A closed queueing network (shown below) comprises two $M/M/1$ nodes, **A** and **B** between which n identical jobs circulate. The nodes have service rates μ_A and μ_B respectively. Upon completion at **A**, a job moves to **B** with probability p and otherwise it remains at **A**. Similarly, upon completion at **B**, a job moves to **A** with probability q.

   Derive a Markov-chain model of this system, explaining what each state in the model signifies and what transition rates exist between states.