

CSCE 317
Spring 2014
Solutions to Second Midterm Exam

1. **(Accept-reject method; 10 points)** Using the accept-reject method, show how to sample the r.v. X whose pdf is

$$f_X(t) = \begin{cases} 12t^2(1-t) & \text{if } 0 \leq t \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

given that you can sample $Y \sim \text{Uniform}(0, 1)$. What is the expected number c of repetitions needed to get a good sample? Show your work.

Answer: We first find c :

$$c = \sup_{0 \leq t \leq 1} \frac{f_X(t)}{f_Y(t)} = \sup_t \frac{f_X(t)}{1} = \sup_t f_X(t) = 12 \sup_t t^2(1-t).$$

Since $f_X(t)$ is differentiable in the interval $(0, 1)$, and its values at $t = 0$ and $t = 1$ are both 0, the supremum is achieved when $(d/dt)(f_X(t)) = 0$. We have

$$\frac{d}{dt} t^2(1-t) = \frac{d}{dt} (t^2 - t^3) = 2t - 3t^2 = 0$$

when $t = 0$ or $t = 2/3$. The larger value is then

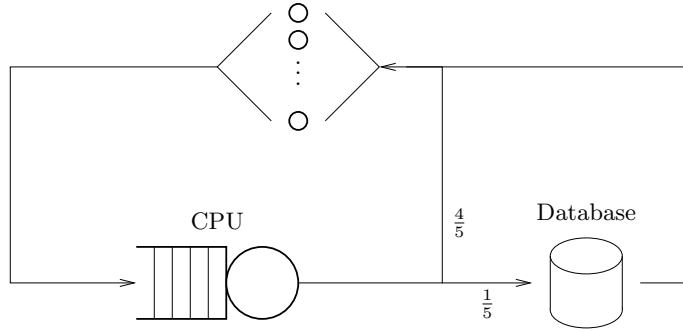
$$f_X(2/3) = \frac{16}{9} = c.$$

The algorithm is thus:

- (a) Generate $u \sim \text{Uniform}(0, 1)$.
 - (b) Accept and output u with probability $f_X(u)/c$:
 - i. Generate $x \sim \text{Uniform}(0, 1)$.
 - ii. If $x \leq (27/4)u^2(1-u)$, then accept and output u .
 - iii. Otherwise, repeat the whole algorithm.
2. **(An ergodic system; 5 points)** Over the course of a day, customers walk into a bank at an average rate of one customer every 20 seconds. There are on average 15 customers in the bank at any given time. How much time does an average customer spend in the bank (in minutes)?

Answer: By Little's law for an open system, we have $\mathbf{E}[N] = \lambda \mathbf{E}[T]$, where N is the number of customers in the bank, λ is the average arrival rate, and T is the time a customer spends in the bank. We are given that $\mathbf{E}[N] = 15$ and that $\lambda = 1/20$ customers per second, or equivalently, $\lambda 60/20 = 3$ customers per minute. We then find $\mathbf{E}[T] = 15/3 = 5$ (minutes).

3. (An interactive system; 15 points) Consider the following interactive system:



Suppose we are given the following information:

- mean user think time = 4 seconds
- expected time spent at the database = 2.5 seconds/visit
- expected number of jobs at the database = 7.5
- expected response time of the system = 2 seconds/job (this does not include think time)
- average CPU usage (service demand) is 0.02 seconds/job

What is the expected number $\mathbf{E}[N_{CPU}^{\text{job}}]$ of jobs queued at the CPU?

Answer: By “job” we mean a system completion. We are given:

$$\begin{aligned} \mathbf{E}[Z] &= 4 \text{ sec} & \mathbf{E}[R] &= 2 \text{ sec/job} \\ \mathbf{E}[T_{DB}] &= 2.5 \text{ sec/visit} & \mathbf{E}[D_{CPU}] &= 0.02 \text{ sec/job} \\ \mathbf{E}[N_{DB}] &= 7.5 \end{aligned}$$

- We also know based on the diagram that each job visits the CPU exactly once, and so $D_{CPU} = V_{CPU} S_{CPU} = S_{CPU}$, i.e., $\mathbf{E}[S_{CPU}] = 0.02 \text{ sec/visit}$.
- Also based on the diagram, we know that the expected throughput at the database $X_{DB} = X_{CPU}/5 = X/5$.
- By Little’s law, $X_{DB} \mathbf{E}[T_{DB}] = \mathbf{E}[N_{DB}] = 7.5$, and so $X_{DB} = 7.5/2.5 = 3 \text{ visits/sec}$.
- Plugging this into the previous equation, we get $X = X_{CPU} = 5X_{DB} = 15 \text{ visits/sec} = 15 \text{ jobs/sec}$.
- We can now use the response time law to find N : We have $2 = \mathbf{E}[R] = N/X - \mathbf{E}[Z] = N/15 - 4$, and so $N = (6)(15) = 90 \text{ jobs in the whole system}$.
- We will need to find the expected number $\mathbf{E}[N_{CPU}]$ of jobs at the CPU. Each of the 90 jobs in the whole system is either at the CPU, at the database, or being thought about by a user, and so we have $90 = \mathbf{E}[N_{CPU}] + \mathbf{E}[N_{DB}] + \mathbf{E}[N_{\text{think}}]$. This gives $\mathbf{E}[N_{CPU}] = 90 - \mathbf{E}[N_{DB}] - \mathbf{E}[N_{\text{think}}] = 82.5 - \mathbf{E}[N_{\text{think}}]$, but we still need $\mathbf{E}[N_{\text{think}}]$.
- We get $\mathbf{E}[N_{\text{think}}]$ by Little’s law applied to the open system of the thinking users: $\mathbf{E}[N_{\text{think}}] = X \cdot \mathbf{E}[T_{\text{think}}] = X \cdot \mathbf{E}[Z] = (15)(4) = 60 \text{ jobs}$.
- Plugging this into the previous equation, we get $\mathbf{E}[N_{CPU}] = 82.5 - 60 = 22.5 \text{ jobs}$.

- The jobs at the CPU are split between those waiting in the queue and those currently running: $22.5 = \mathbf{E}[N_Q^{\text{cpu}}] + \mathbf{E}[N_{\text{running}}]$.
 - If we apply Little's law to the actual processor (not including the queue), we get $\mathbf{E}[N_{\text{running}}] = X_{\text{CPU}} \cdot \mathbf{E}[S_{\text{CPU}}] = (15)(0.02) = 0.3$ jobs. (In this specific case, we are really invoking the utilization law: $\rho_{\text{CPU}} = \lambda_{\text{CPU}}/\mu_{\text{CPU}} = X_{\text{CPU}} \cdot \mathbf{E}[S_{\text{CPU}}] = 0.3$.)
 - Finally, plugging this into the previous equation, we have $\mathbf{E}[N_Q^{\text{cpu}}] = 22.5 - 0.3 = 22.2$ jobs in the queue (on average).
4. **(Modification analysis; 10 points)** The following measurements were obtained over a long stretch of time for an interactive system consisting of a CPU and a slow disk:
- $B_{\text{cpu}} = 250$ seconds
 - $B_{\text{slowdisk}} = 750$ seconds
 - $C = C_{\text{cpu}} = 500$ jobs
 - $C_{\text{slowdisk}} = 1000$ jobs
 - $\mathbf{E}[Z] = 10$ seconds
 - $N = 30$ users
- (a) What is N^* ? Would you say that N is high or low compared with N^* ?
- (b) Which of the following two options would improve the response time more: (i) replacing the CPU with one twice as fast; or (ii) replacing the disk with one twice as fast? Explain.
- (c) Suppose you install another disk twice as fast as the one you currently have, and balance the load between the two disks. Is the CPU now the bottleneck? Explain.

Answer: We observe the following system parameters:

$$D_{\text{cpu}} = \frac{B_{\text{cpu}}}{C} = \frac{250}{500} = 0.5 \text{ sec/job}$$

$$D_{\text{slowdisk}} = \frac{B_{\text{slowdisk}}}{C} = \frac{750}{500} = 1.5 \text{ sec/job} = D_{\text{max}}$$

$$D = D_{\text{cpu}} + D_{\text{slowdisk}} = 2 \text{ sec/job}$$

$$S_{\text{slowdisk}} = \frac{B_{\text{slowdisk}}}{C_{\text{slowdisk}}} = \frac{750}{1000} = 0.75 \text{ sec/visit}$$

(a)

$$N^* = \frac{D + \mathbf{E}[Z]}{D_{\text{max}}} = \frac{2 + 10}{1.5} = 8.$$

So N is high compared with N^* ($30 \gg 8$).

- (b) Replacing the disk would improve the response time more. The disk is the bottleneck device. Replacing the CPU will make hardly any difference, because doing so does not change D_{max} , which controls the response time when N is high.
- (c) Actually, all three devices have the same load in this case, so all three are bottleneck devices, including the CPU. We see this as follows:

- The current average number of disk accesses per job is

$$V_{\text{slowdisk}} = \frac{C_{\text{slowdisk}}}{C} = \frac{1000}{500} = 2 \text{ visits/job} .$$

- After buying the fast disk, we balance the load by finding the new average number of visits $\tilde{V}_{\text{slowdisk}}$ and $\tilde{V}_{\text{fastdisk}}$ to the slow and fast disk, respectively. These quantities should satisfy

$$\tilde{V}_{\text{slowdisk}} + \tilde{V}_{\text{fastdisk}} = 2 \quad (\text{this is the current } V_{\text{slowdisk}})$$

$$S_{\text{fastdisk}} \cdot \tilde{V}_{\text{fastdisk}} = S_{\text{slowdisk}} \cdot \tilde{V}_{\text{slowdisk}} \quad (\text{i.e., } \tilde{D}_{\text{fastdisk}} = \tilde{D}_{\text{slowdisk}})$$

- Since the fast disk is twice as fast as the slow one, we have $S_{\text{slowdisk}} = 2S_{\text{fastdisk}}$, and so the second equation above becomes

$$\tilde{V}_{\text{fastdisk}} = 2\tilde{V}_{\text{slowdisk}} .$$

- Solving the two equations then gives

$$\tilde{V}_{\text{fastdisk}} = \frac{4}{3} , \quad \tilde{V}_{\text{slowdisk}} = \frac{2}{3} .$$

- So we finally get $\tilde{D}_{\text{fastdisk}} = \tilde{D}_{\text{slowdisk}} = S_{\text{slowdisk}} \cdot \tilde{V}_{\text{slowdisk}} = (3/4)(2/3) = 0.5 = D_{\text{cpu}}$.
So all devices have the same load.

(By the way, the new threshold is $\tilde{N}^* = 23$, and so N is still (somewhat) high.)

5. **(Inverse transform method; 10 points extra credit)** Using the inverse transform method, show how to sample the r.v. X whose pdf is

$$f_X(t) = \begin{cases} 2t & \text{if } 0 \leq t \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

given that you can sample $Y \sim \text{Uniform}(0, 1)$. Show your work.

Answer: For $0 \leq x \leq 1$, we have

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_0^x 2t dt = x^2 .$$

So if $F_X(x) = u$, then $x = \sqrt{u}$. So here is the algorithm:

- Generate $u \sim \text{Uniform}(0, 1)$.
- Output \sqrt{u} .