Model solutions + mark scheme for G202/M033

1. I have taken a set of measurements of n predictor variables $w_1, \ldots, w_n$ and n response variables $X_1, \ldots, X_n$. I believe that the response variables are independent, and that $X_i \sim \mathrm{Exp}(\lambda w_i)$, where λ is unknown. Calculate the maximum likelihood estimate for λ.

[11 marks]

4: correct density

density of X_i is $f_i(x) = \lambda w_i e^{-\lambda w_i x}$

density of $X_1, \cdots, X_n$ is $\prod f_i(x_i) = \lambda^n \prod w_i e^{-\lambda \sum w_i x_i}$

$\log \text{lik}(\lambda \mid \underline{x}) = n \log \lambda - \lambda \sum w_i x_i$.

4: what to solve

To find MLE, we want the value of λ that maximizes $\log \text{lik}(\lambda \mid \underline{x})$:

3: execution & answer

$$\frac{d}{d\lambda} = \frac{n}{\lambda} - \sum w_i x_i = 0 \Rightarrow \lambda = \frac{n}{\sum w_i x_i}.$$

2. What is meant by an $M/M/1/B$ queue? Supposing a $M/M/1/B$ queue is currently empty, let t be the expected time until it becomes full. Write down a set of equations that one could solve to find t.

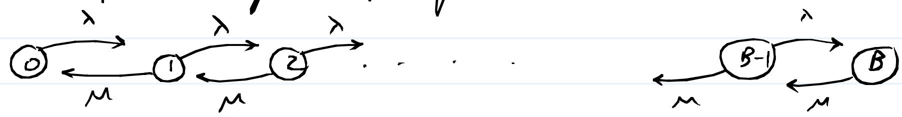
A M/M/1/B queue has

4

- Poisson arrivals, say of rate λ
- Exp. service times, say of rate μ (all independent)
- A single server
- A buffer of size B.
- FCFS discipline

State space diagram for queue size:

2



Let t_i = $\mathbb{E}$ time to hit B starting from i. Then

5: boundary general form additive part

$$t_B = 0$$
$$t_{B-1} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} t_B + \frac{\mu}{\lambda+\mu} t_{B-2}$$
$$\vdots$$
$$t_1 = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} t_2 + \frac{\mu}{\lambda+\mu} t_0$$
$$t_0 = \frac{1}{\lambda} + t_1.$$

use for t

The question asks for $t \equiv t_0$.

3. Find the average number of active calls on an Erlang link with C circuits, arrival rate λ and mean call duration m.

4: correct formula explain terms

The blocking probability on such a link is

$$B = E(\rho, c) = \frac{\rho^c/c!}{\sum_{i=0}^{c} \rho^i/i!} \quad \text{where } \rho = \lambda m.$$

Little's Law says:

4: correct formula explain terms

$$N = \lambda W$$

$$\text{average occupancy of a system} = \text{arrival rate} \times \text{average time spent in the system.}$$

4: correct use identifying λ, W

Here, $N = \underbrace{\lambda(1-B)}_{\text{since this is the rate at which calls actually enter the link}} \times m$

once in, they all have mean duration m.

4. This question concerns a simple slotted-time model for retransmissions by a node in a wireless network.

The node keeps track of the number of transmission failures experienced by the packet it is currently trying to send. In each timeslot, the node may stay silent or it may attempt transmission; if it attempts transmission, it may either succeed or fail. If it succeeds, then it starts over with a new packet in the next timeslot. If it fails, it will retransmit the current packet in a future timeslot.

5: overall clarity & quality of exposition

Mathematically, let X_n be the number of transmission failures experienced by the current packet up to the beginning of timeslot n; then $X_{n+1} = X_n$ if the node stays silent, $X_{n+1} = 0$ if its transmit attempt succeeds, $X_{n+1} = X_n + 1$ if its transmit attempt fails. The probability of attempting transmission in timeslot n is $1/2^{X_n}$. The probability that a transmit attempt fails is q, and failures are assumed to be independent.

(a) Find the distribution of the backoff, i.e. the wait until the next transmission attempt, after a packet has just experienced its kth failure. Show that the average backoff doubles after each transmission failure.

$\mathbb{P}(\text{xmit attempt}) = \frac{1}{2^k}$.

5: identify Geom distribution

So $\mathbb{P}(\text{wait} = 1) = \frac{1}{2^k}$

$\mathbb{P}(\text{wait} = 2) = (1 - \frac{1}{2^k})\frac{1}{2^k}$ etc.

$\text{wait} \sim \text{Geom}(\frac{1}{2^k})$

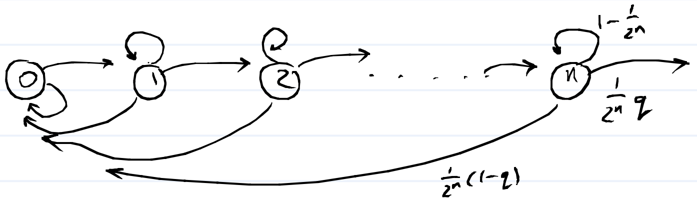
$\mathbb{E}\,\text{wait} = 2^k$.

3: mean interpretation

This clearly doubles whenever k increments.

(b) Draw the state space diagram for the Markov chain X_n. For what values of q is it stable? Calculate the equilibrium distribution. Show that, in equilibrium, the probability of attempting transmission is $p = (1-2q)/(1-q)$.

6: diagram correct states correct arrows correct probs



4: eqm dist balance eqns soln to balance need for normalization correct normalization execution

Balance equations: $\pi_n = \pi_{n-1} \cdot \frac{1}{2^{n-1}} q + \pi_n (1 - \frac{1}{2^n})$

$\Rightarrow \pi_n \cdot \frac{1}{2^n} = \pi_{n-1} \cdot \frac{q}{2^{n-1}} \Rightarrow \pi_n = 2q\,\pi_{n-1}$.

The general solution: $\pi_n = \pi_0 (2q)^n$. For stability, need $q < \frac{1}{2}$.

Then, normalization $\Rightarrow \sum \pi_n = 1 \Rightarrow \frac{\pi_0}{1-2q} = 1 \Rightarrow \pi_0 = 1-2q$. So $\pi_n = (1-2q)(2q)^n$

5: realizing what to do execution

$\mathbb{P}(\text{attempt xmission}) = \sum_n \mathbb{P}(\text{in state } n \;\&\; \text{attempt xmission}) = \sum (1-2q)(2q)^n \cdot \frac{1}{2^n}$

$= (1-2q)\sum q^n = \frac{1-2q}{1-q}$.

5. (a) In a model for Dynamic Alternative Routing of calls on a fully-connected network of links each with C circuits, the equation

$$B = E\big(\lambda/\mu + 2\lambda B(1-B)/\mu, C\big)$$

can be used to compute the probability B that a given link has all its circuits busy. Explain the terms in the equation, and describe how to solve it numerically.

5: overall exposition

2: terms

3: explain parts

This is all bookwork. Let there be n nodes, let calls arrive at rate λ between every pair of nodes, let all holding times be $\mathrm{Exp}(\mu)$. Then,

$$\text{total traffic offered to a given link } L\leftrightarrow M = \lambda + 2(n-2)\lambda\, B\, \tfrac{1}{n-2}\,(1-B)$$

direct traffic for link $L\leftrightarrow M$

number of other node-pairs e.g. $L\leftrightarrow N$ that could use $L\leftrightarrow M$ as part of a two-hop route $L\leftrightarrow M\leftrightarrow N$

P(such a call is blocked on its direct route $L\leftrightarrow M$ and tries a 2-link route)

P(it chooses a 2-hop route that uses link $L\leftrightarrow M$, e.g. $L\leftrightarrow M\leftrightarrow N$)

P(the other leg of this 2-link route does not block the call)

Thus $$B = E\left(\tfrac{1}{\mu}[\lambda + 2\lambda B(1-B)], C\right).$$

3: update method inc. baby steps (here or below)

According to the iterative fixed-point method, we might

- pick an arbitrary starting guess, $B^{(0)}$
- update it by $B^{(n+1)} = E\left(\tfrac{1}{\mu}[\lambda + 2\lambda B^{(n)}(1-B^{(n)})], C\right)$
- keep on updating until $B^{(n)}$ settles down.

We might need to take "baby steps" $B^{(n+1)} = (1-\delta)B^{(n)} + \delta E(\text{—})$ for some small enough δ, to ensure that the fixed-point approach converges.

(b) Consider the following simple slotted-time model for a wireless network consisting of N nodes. In each timeslot, the probability that a node attempts transmission is p. The transmission succeeds if none of the other nodes is also attempting transmission, otherwise it fails. Let the probability of failure be q. A careful analysis of the retransmission mechanism (in Question 4) shows that $p = (1-2q)/(1-q)$.

Write down an expression for q in terms of p, and use this to find a fixed-point equation for q. Compute the numerical solution for q when $N = 10$. Also compute the probability that there is a successful transmission in a given timeslot (i.e. the total throughput of the wireless network).

4: expression for q

$$q = \mathbb{P}(\text{at least one of the others is xmitting}) = 1 - \mathbb{P}(\text{all silent}) = 1-(1-p)^{N-1}.$$

4: suitable form for FP iteration

$$\Rightarrow 1-q = \left(1 - \tfrac{1-2q}{1-q}\right)^{N-1} = \left(\tfrac{1-q-1+2q}{1-q}\right)^{N-1} = \left(\tfrac{q}{1-q}\right)^{N-1} \Rightarrow (1-q)^N = q^{N-1}$$

$$\Rightarrow 1-q = q^{1-\frac{1}{N}} \Rightarrow q = 1-q^{1-\frac{1}{N}}.$$

7: execution

sensible start

technique

spotting if things are going wrong

At $N=10$, a few iterations of $q \leftarrow 1-q^{1-\frac{1}{10}}$ starting at $q=0.4$ converges to 0.482

(For this form of the update eqn, baby steps are needed.)

5: knowing what to calculate

execution

$$\mathbb{P}(\text{succ. xmission}) = \mathbb{P}(\text{exactly 1 xmits}) = Np(1-q) = 37\%.$$

6. In the standard processor-sharing model of TCP, we assume that when a link with speed C is used by n TCP flows then each flow gets throughput $\theta(n) = C/n$. However, cable network engineers have found that, because of inefficiencies in the MAC protocol for cable access links, the achievable utilization is lower when several flows are sharing the link: each flow gets throughput

5: overall exposition

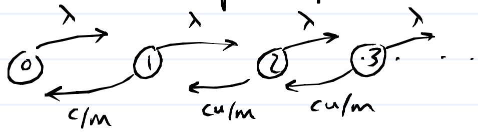
$$\theta(n) = \begin{cases} C & \text{if } n = 1 \\ Cu/n & \text{if } n > 1 \end{cases}$$

for some constant $0 < u < 1$.

(a) Calculate the average flow completion time, as a function of the arrival rate λ, the mean job size m, the link speed C and the parameter u. You may find useful the formulae on the next page.

Markov chain for # flows active: by considering residual file sizes,

3: statespace diagram



First we use detailed balance to find eqm distribution:

3: detailed balance

$$\left.\begin{aligned} \pi_0 \lambda &= \pi_1 \frac{C}{m} \\ \pi_1 \lambda &= \pi_2 \frac{Cu}{m} \\ &\text{etc.} \end{aligned}\right\} \Rightarrow \quad \pi_1 = \frac{\lambda m}{C}\pi_0, \quad \pi_2 = \frac{\lambda m}{Cu}\frac{\lambda m}{C}\pi_0$$

$$\pi_3 = \left(\frac{\lambda m}{Cu}\right)^2 \frac{\lambda m}{C}\pi_0 \text{ etc.}$$

2: idea of normalization

Normalization: $\pi_0 \left[1 + \frac{\lambda m}{C} + \frac{\lambda m}{C}\frac{\lambda m}{Cu} + \frac{\lambda m}{C}\left(\frac{\lambda m}{Cu}\right)^2 + \cdots \right] = 1$

3: execution

$$\Rightarrow \pi_0 \left(1 + \frac{\frac{\lambda m}{C}}{1 - \frac{\lambda m}{Cu}} \right) = 1 \Rightarrow \pi_0 = \frac{1-\rho}{1+\rho^2\left(\frac{1}{u}-1\right)} \quad \text{where } \rho = \frac{\lambda m}{C}.$$

4: what to do with # flows

$$\text{Av. \# flows} = \sum n\pi_n = \pi_0\left(\rho + \frac{2\rho^2}{u} + \frac{3\rho^3}{u^2} + \cdots\right) = \pi_0 u \left(\frac{\rho}{u} + 2\frac{\rho^2}{u^2} + \cdots\right) = \pi_0 u \frac{\rho/u}{(1-\rho/u)^2}$$

$$= \frac{\rho(1-\rho)}{(1+\rho^2(\frac{1}{u}-1))(1-\rho/u)^2}. \qquad \text{Sanity check: if } u=1, \text{ get } \frac{\rho}{1-\rho}.$$

3: use of Little

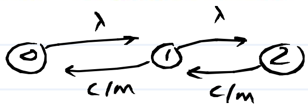
1: correct answer!

By Little's law, average wait $= \frac{1}{\lambda} \times$ av. # flows.

(b) Calculate the average flow completion time under FCFS discipline.

2: diagram

4: recognition & answer.

Markov chain:  just like standard processor sharing, for which average # flows $= \frac{\rho}{1-\rho}$, average wait $= \frac{1}{\lambda}\frac{\rho}{1-\rho}$.

(c) It can be shown that the average flow completion time is smaller under FCFS. Explain briefly how FCFS might nevertheless lead to *worse* quality of service for some users.

3: sensible remarks

Problem: a huge flow will block others for a long time, eg. P2P upload.
(NB The formula for flow completion time under FCFS only applies when job sizes are $\text{Exp}(\frac{1}{m})$, not when some jobs are big & others much smaller.)

7. The MAC protocol for a cable access uplink works as follows. Time is slotted, and each timeslot is divided into a request subslot, a reply subslot, and a data subslot. In the request subslot, users with packets to send can signal a request to the cable head-end (CHE). The CHE chooses one of these users, and in the reply subslot the CHE signals its choice. In the data subslot, the chosen user transmits his packet. However, request signals can interfere with each other, and when this happens the timeslot is wasted and users have to wait until the next timeslot to re-signal their requests. To avoid this problem, the MAC protocol has an additional request mechanism: the chosen user in timeslot n can append a request for timeslot $n+1$ to his packet transmission in timeslot n, and the CHE will grant it. This has the effect of giving priority to ongoing flows. Of course, it only works if the chosen user has another packet ready to send.

5: overall exposition

We may model this as follows. Let x be the throughput of a user, in packets per second. There is some minimum throughput $x_{\min}$ such that if $x > x_{\min}$ then the priority mechanism works and the user experiences round trip time RTT_0. If $x < x_{\min}$ then there is not enough data for the priority mechanism to work, so the user experiences round trip time $\mathrm{RTT}_0 + \Delta$, for some constant $\Delta > 0$.

Write down a drift model for the throughput of a TCP flow using this cable access link. Sketch the drift diagram. Find any fixed points. Describe the behaviour that this drift model predicts.

5: sensible drift; realise the role of RTT.

let x_t = throughput at time t, w_t = window size at time t.

$$\frac{dw_t}{dt} = \frac{1}{RTT(x_t)} - \frac{p\, w_t^2}{2\, RTT(x_t)} \qquad \text{(reasoning as in lecture notes)}$$

3: drift for x rather than w.

$$\Rightarrow \frac{dx_t}{dt} = \frac{1}{RTT(x_t)^2} - p\frac{x_t^2}{2}.$$

To find fixed points, we need to solve

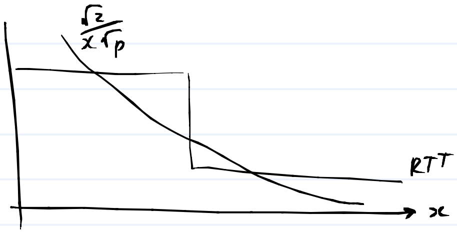
4: idea of FP soln

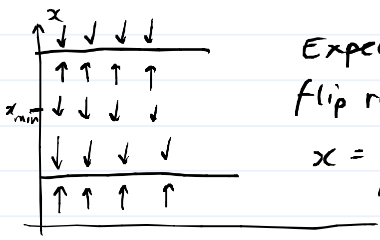
$$\frac{1}{RTT(x)^2} - \frac{p x^2}{2} = 0 \Rightarrow RTT(x) = \frac{\sqrt{2}}{x\sqrt{p}}.$$

There are three cases:

4: realize & argue separate cases

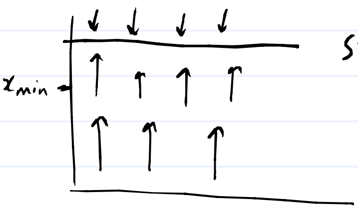
4 correct identification of cases





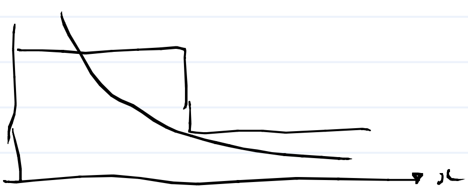
Expect bistable behaviour: flip randomly from $x = \frac{\sqrt{2}}{RTT_0\sqrt{p}}$ to $x = \frac{\sqrt{2}}{(RTT_0+\Delta)\sqrt{p}}$

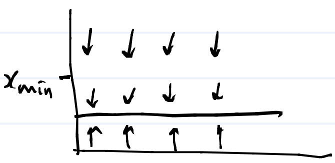
4: identify fixed pts.



Stay close to $x = \frac{\sqrt{2}}{RTT_0\sqrt{p}}$

4: qualitative description of behaviour inc. bistability





Stay close to $x = \frac{\sqrt{2}}{(RTT_0+\Delta)\sqrt{p}}$.

The low-throughput fixed pt is seen if $\frac{\sqrt{2}}{(RTT_0+\Delta)\sqrt{p}} < x_{min}$.

The high-throughput fixed pt is seen if $\frac{\sqrt{2}}{RTT_0\sqrt{p}} > x_{min}$