IEOR 6711, Solutions to HMWK 5, Professor Sigman

1. Consider a FIFO $M/M/1$ queue (arrival rate $0 < \lambda < \infty$, service rate $0 < \mu < \infty$; $\rho = \lambda/\mu$). Let $X(t)$ denote the number of customers in the system at time t, a birth and death process.

   (a) Suppose that $\rho > 1$. Let $\{X_n : n \geq 0\}$ denote the embedded discrete-time Markov chain, and suppose that $X_0 = 1$. What is the probability that this chain will go off to ∞ without ever going to state 0 first?
   SOLUTION: The embedded chain is a simple random walk (restricted to be non-negative): $P_{0,1} = 1$ and $P_{i,i+1} = p$, $P_{i,i-1} = q = 1 - p$, where $p = \lambda/(\lambda + \mu) > 0.5$ since $\rho > 1$. Thus from the gambler's ruin problem, the answer is given by $P_1(\infty) = 1 - (q/p) = 1 - \rho^{-1} > 0$.

   (b) *Continuation:* Use (a) to prove that $\{X(t)\}$ is transient when $\rho > 1$.
   SOLUTION: (It suffices to prove that state 0 is transient since the chain is irreducible.) If we start in state 0, then since $P_{0,1} = 1$, the above computation yields $\rho^{-1} = f_0 =$ the probability the chain will ever return to 0 given $X_0 = 0$. $f_0 < 1$, so we conclude, by definition, that state 0 is transient.

   (c) Suppose now that $\rho = 1$. Prove that $\{X(t)\}$ is null recurrent.
   SOLUTION: From birth and death CTMC general results (balance equations), we know that the continuous-time chain is *positive recurrent* if and only if $\rho < 1$. Thus, when $\rho = 1$ it is either transient or null recurrent. But if it were transient, then so would be the embedded chain, but the embedded chain is recurrent because when $\rho = 1$, $p = 0.5$ for the random walk; the simple symmetric case.

2. Consider a stable FIFO $M/M/1$ queue, $0 < \rho < 1$. Let $X(t)$ denote the number in system at time t, and let $P_n = (1 - \rho)\rho^n$, $n \geq 0$ denote the stationary distribution. Show (direct calculation) that if $X(0) \sim (P_n)$ (e.g., the chain is started off with its stationary distribution, hence is a stationary process), then the time until the first departure (after time $t = 0$), t_1^d, has an exponential distribution at rate λ.

   SOLUTION: We condition on wether the server is busy or not at time $t = 0$. If busy (probability $\rho = 1 - P_0$), then the next departure will occur after the (remaining) service completion of length S that is exponential at rate μ by the memoryless property; $t_1^d = S \sim exp(\mu)$. If not busy (probability $1 - \rho = P_0$), then first an arrival must come (remaining length $T \sim exp(\lambda)$ by the memoryless property), and then this customer must enter service for an amount of time $S \sim exp(\mu)$; so in this case $t_1^d = T + S \sim \exp(\lambda) \star \exp(\mu)$ (a convolution).

   We now can compute the Laplace transform of t_1^d yielding (after some algebra) that for any $s \geq 0$,

$$E^{-st_1^d} = \rho\frac{\mu}{\mu + s} + (1 - \rho)\frac{\lambda}{\lambda + s} \times \frac{\mu}{\mu + s} = \frac{\lambda}{\lambda + s},$$

   and indeed $t_1^d \sim exp(\lambda)$.

3. Consider the M/M/1 queue (arrival rate λ service time rate μ) with the following twist: Each customer independently will get impatient after an amount of time that is exponentially distributed at rate γ while waiting in line (queue) and leave before ever entering service, and without ever returning. A customer who does enter service completes service (e.g., customers are only impatient while waiting in the line, not when in service.)

   (a) You arrive finding exactly one customer in the system (hence they are in service) and you join the queue to wait. What is the probability that you will get served?

   SOLUTION: Letting C denote your independent $exp(\gamma)$ rv, independent of the $exp(\mu)$ remaining service time S, we want $P(C > S) = \mu/(\mu + \gamma)$.

   (b) You arrive finding exactly two (2) customers in the system (one in service, one in line)) and you join the end of the queue to wait. What is the probability that you will get served?

   SOLUTION Let S denote the remaining service time of who is in service, and let S_1 denote the service time of the customer waiting in front of you (and let C_1 denote their exponential impatience time, and let C_2 denote yours.) Let D denote the length of time until the server will be free to serve you:

$$D = S + S_1 I\{C_1 > S\}.$$

   We need to compute $P(D < C_2)$ and we do so by conditioning on wether $\{C_1 > S\}$ or $\{C_1 \leq S\}$.

   Given that $S < C_1$, $S = \min\{S, C_1\} \sim exp(\gamma + \mu)$. So letting Y denote an independent exponential at rate $\gamma + \mu$, we have

$$P(D < C_2 \mid S < C_1) = P(Y + S_1 < C_2),$$

   where all three (Y, S_1, C_2) are independent. Meanwhile, if $C_1 < S$, then $D = S$ and

$$P(D < C_2 \mid C_1 < S) = P(S < C_2 \mid C_1 < S) = \frac{P(C_1 < S < C_2)}{P(C_1 < S)}.$$

   Thus unconditioning on the two events yields

$$P(D < C_2) = P(Y + S_1 < C_2)P(S < C_1) + P(C_1 < S < C_2), \qquad (1)$$

   where we now need to compute both $P(Y + S_1 < C_2)$[1] and $P(C_1 < S < C_2)$. $P(Y + S_1 < C_2 \mid Y + S_1 = x) = e^{-\gamma x}$ and hence $P(Y + S_1 < C_2 \mid Y + S_1) = e^{-\gamma(Y+S_1)}$ yielding

$$P(Y + S_1 < C_2) = E[e^{-\gamma(Y+S_1)}] = E[e^{-\gamma Y}]E[e^{-\gamma S_1}].$$

[1] We can also derive $P(Y + S_1 < C_2)$ as follows: First we must have $C_2 > Y$, and then given that, the remainder $C_2 - Y$ must satisfy $C_2 - Y > S_1$. But given $C_2 > Y$, that remainder is yet again like a new independent C_2, thus the answer is the product $P(C_2 > Y)P(C_2 > S_1)$.

But these are just the Laplace transforms of exponentials, evaluated at $s = \gamma$:

$$E[e^{-\gamma Y}] = \frac{\gamma + \mu}{\gamma + \mu + \gamma} = \frac{\gamma + \mu}{2\gamma + \mu}$$

$$E[e^{-\gamma S_1}] = \frac{\mu}{\mu + \gamma};$$

thus

$$P(Y + S_1 < C_2) = \frac{\gamma + \mu}{2\gamma + \mu} \times \frac{\mu}{\mu + \gamma} = \frac{\mu}{\mu + 2\gamma}.$$

$P(C_1 < S < C_2 \mid S = x) = (1 - e^{-\gamma x})(e^{-\gamma x})$ and so (via Laplace transforms again)

$$P(C_1 < S < C_2) = E[e^{-\gamma S}] - E[e^{-2\gamma S}] = \frac{\mu}{\mu + \gamma} - \frac{\mu}{\mu + 2\gamma}.$$

Thus from (1) (and algebra)

$$P(D < C_2) = \frac{\mu}{\mu + 2\gamma}.$$

(It is interesting to see how simple the final expression is; it is the same as $P(S < \min\{C_1, C_2\})$ but I am not sure why. Maybe there is an easier way to solve this problem.)

(c) Model as a Birth and Death process, give the birth and death rates, $\lambda_n, \mu_n,\ n \geq 0$.

SOLUTION

$X(t) =$ the number of customers in the system at time t forms a B& D process. $\lambda_n = \lambda,\ n \geq 0,\ \mu_n = (n-1)\gamma + \mu,\ n \geq 1.\ (\mu_0 = 0)$.

(d) Set up the Birth and Death balance equations for the limiting probabilities P_n (but do not try to solve.)

SOLUTION

$\lambda P_n = (n\gamma + \mu)P_{n+1},\ n \geq 0.$

(e) Compute the ratio P_{n+1}/P_n and prove using the "ratio test" from calculus, that the limiting probabilities exist for all values of $\lambda > 0, \mu > 0, \gamma > 0$. Thus this chain is always positive recurrent (e.g., a condition such as $\rho < 1$ is not needed); explain intuitively why this should be so.

SOLUTION

$$P_{n+1}/P_n = \frac{\lambda}{n\gamma + \mu} \to 0,$$

as $n \to \infty$. In particular the ratio is eventually strictly bounded less than a constant less than 1; hence $\sum_n P_n < \infty$.

Intuitively, the system is always stable because if the line gets too big, then more and more customers will get impatient and leave, hence reducing the congestion.

4. Cars arrive to a parking lot according to a Poisson process at rate λ. Each car, when parking, independently remains parked for an amount of time that is iid with an exponential distribution at rate μ. The parking lot only has c spots, and any car who arrives finding all c spots taken immediately goes nearby to a huge (infinite capacity) lot, and parks there instead (same iid exponential parking times). What is the long-run average number of parked cars in the infinite capacity lot?

   SOLUTION Let $X_1(t)$ denote the number of cars at the first facility, and $X_2(t)$ at the second. Note that $X(t) = X_1(t) + X_2(t)$ is in fact the number in a $M/M/\infty$ queue hence has time-average $l = \rho = \lambda/\mu$. Meanwhile $X_1(t)$ is simply a $M/M/c$ loss queue, hence has stationary probabilities (solution to the balance equations) $P_n = \rho^n/n!P_0,\ 0 \le n \le c$, where

   $$P_0 = \Big[\sum_{n=0}^{c} \rho^n/n!\Big]^{-1}.$$

   So $X_1(t)$ has a time-average given by the mean of this distribution $l_1 = \sum_{n=0}^{c} nP_n$. Thus $X_2(t)$ has time-average $l_2 = l - l_1 = \rho - l_1$, and thus can be computed exactly.

   Alternatively: From Pasta, P_c = the proportion of arrivals to the first facility who are "lost" and thus must go to the second; thus λP_c is the arrival rate to the second facility. From $l = \lambda w$ applied to the second facility, we conclude that $l_2 = (\lambda P_c)(1/\mu) = \rho P_c$.

5. For the M/G/∞ queue, with $L(t)$ denoting number in system at time t with $L(0) = 0$, Let $S_r(t)$ denote the remaining service time at time t of the customer in service, *conditional on the event* $\{L(t) = 1\}$.

   Compute $\lim_{t\to\infty} P(S_r(t) > x | X(t) = 1)$.

   SOLUTION:

   The key idea is to realize that we can treat this customer as having arrived at time $U \sim unif(0,t)$: For if so then $Y(t) = S - (t - U)$, and we want the limit as $t \to \infty$ of

   $$P(S - (t - U) > x | S - (t - U) > 0) = P(S - U > x | S - U > 0),$$

   because $t - U$ itself is Uniform $(0,t)$ if U is. Expanding yields

   $$\begin{aligned}\frac{P(S > x + U)}{P(S > U)} &= \frac{\frac{1}{t}\int_0^t P(S > x + u)du}{\frac{1}{t}\int_0^t P(S > u)du} \\ &= \frac{\int_x^t P(S > u)du}{\int_0^t P(S > u)du} \\ &\to \mu \int_x^\infty P(S > u)du,\end{aligned}$$

   where $\mu = 1/E(S)$.

   This defines the tail of a distribution called the equilibrium distribution of S, which represents the stationary distribution of a remaining service time. The idea is that if way out in the future, at a random time, you found someone in service, then their remaining service time would have this distribution.

To make our proof rigorous we partition $N(t)$ into 3 classes

(1) X_1 = those who are in the system at t with remaining service time $> x$.

(2) X_2 = those who are in the system at t but remaining service time $\leq x$.

(3) X_3 = those who have departed by t.

We know (partitioning theorem) that these three are independent Poisson rvs with means given by

$\alpha_1(t) = \lambda t P_1(t) = \lambda t P(S > U + x) = \lambda \int_x^t P(S > u)du$

$\alpha_2(t) = \lambda t P_2(t) = \lambda t P(U < S \leq U + x) = \lambda \int_0^x P(S > u)du$

$\alpha_3(t) = \lambda t P_3(t) = \lambda t P(S \leq U)$

(Note that $X_1(t) + X_2(t) = X(t)$ is Poisson with mean $\lambda \int_0^t P(S > u)du$.)

$$
\begin{aligned}
P(Y(t) > x | X(t) = 1) &= P(Y(t) > x | X_1(t) + X_2(t) = 1) \\
&= \frac{P(Y(t) > x, X_1(t) + X_2(t) = 1)}{P(X_1(t) + X_2(t) = 1)} \\
&= \frac{P(X_1(t) = 1, X_2(t) = 0)}{P(X_1(t) + X_2(t) = 1)} \\
&= \frac{P(X_1(t) = 1)P(X_2(t) = 0)}{P(X_1(t) + X_2(t) = 1)} \\
&= \frac{\alpha_1(t)}{\alpha_1(t) + \alpha_2(t)} \\
&= \frac{\int_x^t P(S > u)du}{\int_0^t P(S > u)du},
\end{aligned}
$$

and the proof proceeds as before.

6. Suppose that $\{X(t)\}$ is an irreducible (non-explosive) CTMC with transition rates matrix $Q = (q_{i,j})$. Independently suppose that $\{t_n : n \geq 1\}$ is a Poisson process at rate λ.

   (a) Let $Z_n = X(t_n),\ n \geq 1,\ Z_0 = X(0)$. Argue that $\{Z_n : n \geq 0\}$ is an irreducible discrete-time Markov chain with transition matrix $\tilde{P} = (\tilde{P}_{i,j})$ given by

   $$\tilde{P} = [I - (Q/\lambda)]^{-1}.$$

   SOLUTION:
   By independence, each t_n is a stopping time, and hence (via the strong Markov property) given Z_n, $\{X(t_n + t) : t > 0\}$ is yet again the same CTMC but with initial condition Z_n, independent of the past. This will be so for *any* independent renewal process used for sampling. If we let $T_n = t_{n+1} - t_n$

denote the iid interarrival times for the renewal process, then given Z_n we only need the independent (and independent of the past) rv T_n to predict the future value Z_{n+1}.

$\tilde{P}_{i,j} = P(Z_{n+1} = j \mid Z_n = i)$ does not depend on n because the $T_n,\ n \geq 0$ are iid.

Let $T = t_1 \sim exp(\lambda)$, it has density $\lambda e^{-\lambda t},\ t \geq 0$. We know that $P(t) = e^{Qt},\ t \geq 0$. Thus, conditional on $T = t$, $E(\tilde{P} \mid T = t) = P(t) = e^{Qt}$, and so $E(\tilde{P} \mid T) = e^{QT}$. Taking expected values then yields

$$
\begin{aligned}
\tilde{P} &= E(P(T)) && (2)\\
&= \int_0^\infty P(t)\lambda e^{-\lambda t} dt && (3)\\
&= \int_0^\infty P(u/\lambda)e^{-u} du && (4)\\
&= \int_0^\infty P(u/\lambda)e^{-uI} du && (5)\\
&= \int_0^\infty e^{Qu/\lambda}e^{-uI} du && (6)\\
&= \int_0^\infty e^{-u(I-(Q/\lambda))} du && (7)\\
&= [I - (Q/\lambda)]^{-1}. && (8)
\end{aligned}
$$

Irreducibility follows since in fact $\tilde{P}_{i,j} > 0$ for all pairs (i, j), a condition known as *strong irreducibility*: We know that $P_{i,j}(t) > 0$ for some t and since holding times are continuous (exponential) it follows that $P_{i,j}(t) > 0$ within a time interval $t \in (s_1, s_2),\ 0 \leq s_1 < s_2$, and hence

$$\tilde{P}_{i,j} = \int_0^\infty P_{i,j}(t)\lambda e^{-\lambda t} dt \geq \int_{s_1}^{s_2} P_{i,j}(t)\lambda e^{-\lambda t} dt > 0.$$

(b) Prove that $\{X(t)\}$ is positive recurrent if and only if $\{Z_n\}$ is positive recurrent in which case $\pi = \vec{P}$; they share the same stationary distribution.

SOLUTION: By irreducibility, it suffices to prove that a probability solution to $\pi = \pi\tilde{P}$ (for $\{Z_n\}$) exists if and only if a probability solution to $\vec{P}Q = \vec{0}$ (for $\{X(t)\}$) exists and that $\pi = \vec{P}$.

This is immediate: $\pi = \pi\tilde{P}$ if and only if $\pi = \pi[I - (Q/\lambda)]^{-1}$ if and only if $\pi[I - (Q/\lambda)] = \pi$ if and only if $\pi Q = \vec{0}$.

7. Just as in discrete time, if we consider a positive recurrent CTMC in stationarity that has been started since the infinite past, $\{X^*(t) : -\infty < t < \infty\}$, then the (stationary) time-reversal $X^{(r)}(t) = X^*(-t) : t \geq 0$ is itself a CTMC. It has the same holding time rates $\{a_i\}$ and the same stationary distribution $\vec{P} = (P_j)$ as the

original forward time CTMC. The only thing that can differ are the transition rate matrices Q and $Q^{(r)}$.

A positive recurrent CTMC is called *time-reversible* if the time-reversed process $\{X^{(r)} : t \geq 0\}$ has the same distribution as the forward-time process $\{X^*(t) : t \geq 0\}$. This is equivalent to saying that $Q = Q^{(r)}$, which can be stated in terms of the forward-time chain:

> *the long-run rate that the chain moves from i to j equals the long-run rate that the chain moves from from j to i, for any two states $i, j \in \mathcal{S}$.*

Thus a positive recurrent CTMC is time-reversible if for all pairs of states $i, j \in \mathcal{S}$,

$$a_i P_i P_{i,j} = a_j P_j P_{j,i} \tag{9}$$

Since a birth and death process can only make transitions of magnitude ± 1 we see that in such a case Equation (9) reduces to "the long-run rate that the chain moves from i to $i+1$ equals the long-run rate that the chain moves from from $i+1$ to i, for all states $i \in \mathcal{S}$"; the birth and death balance equations. We conclude:

> *Every positive recurrent birth and death (B&D) process is time-reversible.*

(a) *M/M/1 queue:* $X(t) =$ the number of customers in an M/M/1 queue is a B& D process, so we conclude that when $\rho < 1$, it is time-reversible. Assume that it is started at time $t = 0$ with its stationary distribution. Let $\psi = \{t_n : n \geq 1\}$ denote the Poisson arrival times starting from time $t = 0$: $0 < t_1 < t_2 < \cdots$, and let $\psi^{(d)} = \{t_n^d : n \geq 1\}$ denote the point process of departure times after time $t = 0$: $0 < t_1^d < t_2^d < \cdots$. We know from Exercise 2 above that t_1^d has an exponential distribution at rate λ, but here you will deduce more.

Argue (from time-reversibility) that $\psi^{(d)}$ must have the same distribution (as a point process) as ψ, and hence must itself be a Poisson process at rate λ: *the departure process from a stationary M/M/1 queue is itself a Poisson process.*

SOLUTION:

The points of ψ are precisely the times at which the forward process jumps up by 1. Thus they must have the same distribution (from time-reversibility) as the points in time for which the time-reversed process jumps up by 1; but such times have the same distribution as the departure times $\psi^{(d)}$.

(b) Consider an $M/M/\infty$ queue (arrival rate λ service rate μ) and suppose that departures from it immediately attend another facility: a FIFO single-server queue with its own iid exponential service times at rate μ_2. Assuming that $\lambda < \mu_2$, find the long-run average number of customers in the second facility.

SOLUTION: Just as for the M/M/1 queue, $L(t) =$ the number in system for the $M/M/\infty$ model forms a B& D process hence is time-reversible. Thus, in stationarity, its departure process is a Poisson process at rate λ. Feeding that to the second facility thus yields a standard M/M/1 queue with $\rho = \lambda/\mu_2$, and hence average number in system $\rho/(1-\rho)$.

8. Consider a CTMC $\{X(t)\}$ (with $P_{i,i} = 0,\ i \in \mathcal{S}$) with embedded chain transition matrix $P = (P_{i,j})$ and holding time rates $\{a_i\}$. Assume that $a = \sup\{a_i : i \in \mathcal{S}\} <$

∞. Consider an alternative CTMC $\{\overline{X}(t)\}$for which all holding time rates are fixed at the constant rate a; $\overline{a}_i = a$, and has embedded transition probabilities given by

$$\overline{P}_{i,j} = \begin{cases} \frac{a_i}{a} P_{i,j} & \text{if } j \neq i, \\ 1 - \frac{a_i}{a} & \text{if } j = i. \end{cases}$$

(So $\overline{P}_{i,i} > 0$ is possible.)

(a) Let $\overline{N}(t)$ denote the number of transitions by time t for $\{\overline{X}(t)\}$. Explain why $\{\overline{N}(t) : t \geq 0\}$ forms a Poisson process at rate a.

SOLUTION: By the Markov property, for any state i, given $\overline{X}(t) = i$, the chain will spend an exponential amount of time at rate a in state i independent of the past and then move. But this rate a does not even depend on i, so the sequence of consecutive holding times always forms an iid sequence of rvs distributed as exponential a; a Poisson process at rate a.

(b) Show that the balance equations are the same for the two chains.

SOLUTION:

Balance equations for $\overline{X}(t)$ are

$$aP_j = aP_j(1 - \frac{a_j}{a}) + \sum_{i \neq j} P_i a(\frac{a_i}{a} P_{i,j}),\ j \in \mathcal{S}, \tag{10}$$

which algebraically reduce to

$$a_j P_j = \sum_{i \neq j} P_i a_i P_{ij},\ j \in \mathcal{S}. \tag{11}$$

(c) Explain why $\{X(t)\}$ and $\{\overline{X}(t)\}$ have the same distribution as stochastic processes; $P_{i,j}(t) = \overline{P}_{i,j}(t),\ t \geq 0$ for all pairs i, j.

SOLUTION: Recall that "a geometric sum of iid exponentials is yet again exponential". So all that is happening here is that each original holding time $H_i \sim exp(a_i)$ has been broken down and re-expressed as a geometric sum of iid exponentials at rate a in which the geometric has "success" probability a_i/a. That is what the $\overline{P}_{i,j}$ do. So when state i is entered, in both models the *total* amount of time spent there until changing to a state $j \neq i$ is exponential at rate a_i.