Problem Set 9 - Solutions
Due: April 27, 2005

1.

(a) First note that spam messages, invitations and other e-mail are all independent Poisson processes, at rates $p\lambda$, $q\lambda$, and $(1-p-q)\lambda$. The event of the time T at which you decide about your popularity being greater than some value t is the same as the event of having less than 10 messages in each of the IMPORTANT and OTHER folders at time t.

Since arrivals into the IMPORTANT and OTHER folders are independent, the probability that neither has had 10 arrivals by time t is the product of the probabilities of each receiving less than 10 by that time. Since for each, the number of messages received in $[0, t]$ is a Poisson random variable, we have

$$\begin{aligned}\mathbf{P}(T > t) &= \sum_{i=0}^{9} \frac{(\lambda q t)^i e^{-\lambda q t}}{i!} \sum_{j=0}^{9} \frac{(\lambda(1-p-q)t)^j e^{-\lambda(1-p-q)t}}{j!} \\ &= e^{-\lambda(1-p)t} \sum_{i=0}^{9} \frac{(\lambda q t)^i}{i!} \sum_{j=0}^{9} \frac{(\lambda(1-p-q)t)^j}{j!}\end{aligned}$$

and the CDF is

$$F_T(t) = 1 - \mathbf{P}(T > t).$$

(b) At least 10 of the first 19 non-spam messages must be party invitations. Note that if I receive my 10th invitation by the time I get the 19th non-spam message, since there will be fewer than 10 messages collected in the OTHER folder by that time, I will conclude that I am very popular. Conversely, if I had fewer than 10 items in the INVITATION folder by the time I have received 19 non-spam messages, then I must have at least 10 in the OTHER folder and would conclude that I am not very popular

(c) Having showed in part (b) that getting at least 10 invitations among the first 19 non-spam messages is necessary and sufficient to conclude "very popular" overall, we are looking for the probability of that event. The number of invitations in the first 19 non-spam messages is a binomial random variable, since each non-spam message is independently an invitation with probability $q/(1-p)$. So, the probability of concluding "very popular" is

$$\mathbf{P}(\text{very popular}) = \sum_{i=10}^{19} \binom{19}{i} (q/(1-p))^i ((1-p-q)/(1-p))^{19-i}.$$

Here is an alternative solution: The probability that I conclude that I am very popular upon receiving kth non-spam message, $10 \le k \le 19$, is the probability that exactly 9 out of the first $k-1$ are invitations and and the kth is also an invitation, i.e., $\binom{k-1}{9} (q/(1-p))^{10}((1-p-q)/(1-p))^{k-10}$. Summing over k,

$$\mathbf{P}(\text{very popular}) = \sum_{k=10}^{19} \binom{k-1}{9} (q/(1-p))^{10}((1-p-q)/(1-p))^{k-10}.$$

It can be shown that this is equal to the first answer found above.

2. (a) We have modeled the fouls as a Poisson process with parameter $\lambda = \frac{1}{8}$. This implies that, neglecting fouling out, the number of fouls in the interval $[0,t]$ is a Poisson random variable with parameter $\lambda t = \frac{1}{8}t$. All we have to do is adjust for the fact that more than 6 fouls cannot be commited by assigning the probabilities of 7, 8, ... fouls to 6 to obtain:

$$p_{X_t}(k) = \begin{cases} e^{-t/8}\dfrac{(t/8)^k}{k!}, & \text{for } k = 0, 1, \ldots, 5; \\ \displaystyle\sum_{\ell=6}^{\infty} e^{-t/8}\dfrac{(t/8)^\ell}{\ell!}, & \text{for } k = 6, \end{cases}$$
$$= \begin{cases} e^{-t/8}\dfrac{(t/8)^k}{k!}, & \text{for } k = 0, 1, \ldots, 5; \\ 1 - \displaystyle\sum_{\ell=0}^{5} e^{-t/8}\dfrac{(t/8)^\ell}{\ell!}, & \text{for } k = 6. \end{cases}$$

(b) Wallace has fouled out if at the end of the game he has six fouls. Thus the probability of fouling out is

$$p_{X_{48}}(6) = 1 - \sum_{\ell=0}^{5} e^{-48/8}\frac{(48/8)^\ell}{\ell!}$$
$$= 1 - e^{-6} - 6\,e^{-6} - \frac{1}{2}6^2\,e^{-6} - \frac{1}{6}6^3\,e^{-6} - \frac{1}{24}6^4\,e^{-6} - \frac{1}{120}6^5\,e^{-6} \approx 0.5543.$$

(c) When Wallace does not foul out, his time wins by (0.25 points/minute)·(48 minutes) = 12 points. When Wallace fouls out at time t, the score differential in favor of Wallace's team is

$$\underbrace{0.25t}_{\text{Wallace playing}} - \underbrace{0.5(48-t)}_{\text{Wallace not playing}} - \underbrace{6}_{\text{Wallace's three technical fouls}} = 0.75t - 30.$$

Thus Wallace's team wins whenever $0.75t - 30 > 0$ or $t > 40$.

In terms of our model for Wallace's fouls, we are interested in the sixth arrival time. Without restricting the game length to 48 minutes for the moment, this sixth arrival time has the Erlang PDF of order 6 and parameter $\lambda = \frac{1}{8}$. Denoting the sixth arrival time by U, we have

$$f_U(u) = \frac{\lambda^6 u^5 e^{-\lambda u}}{5!}, \qquad u \geq 0.$$

The event of interest is $\{U > 40\}$ because this captures Wallace not fouling out and Wallace fouling out late enough in the game that his team wins.

$$\mathbf{P}(\{\text{Wallace's team wins}\}) = \mathbf{P}(U > 40) = \int_{40}^{\infty} f_U(u)\,du = \int_{40}^{\infty}\frac{\lambda^6 u^5 e^{-\lambda u}}{5!}\,du$$
$$= \left(\frac{1}{8}\right)^6 \frac{1}{5!}\int_{40}^{\infty} u^5 e^{-\frac{1}{8}u}\,du$$
$$= \frac{1097}{12}e^{-5} \approx 0.616$$

where the integral can be computed by repeated integration by parts.

(d) This is very similar to the previous part. When Wallace commits his fifth foul at time t, the score differential in favor of Wallace's team is

$$\underbrace{0.25t}_{\text{Wallace playing}} - \underbrace{0.5(48-t)}_{\text{Wallace not playing}} = 0.75t - 24.$$

Thus Wallace's team wins whenever $0.75t - 24 > 0$ or $t > 32$.

We are now interested in the fifth arrival time in the Wallace-foul process. This fifth arrival time has the Erlang PDF of order 5 and parameter $\lambda = \frac{1}{8}$. Denoting the fifth arrival time by V, we have

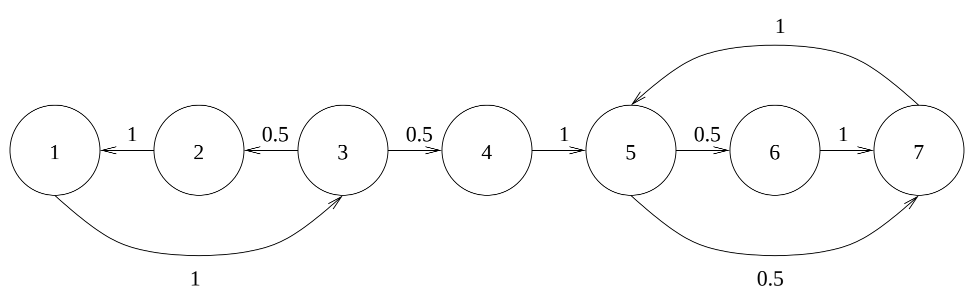
$$f_V(v) = \frac{\lambda^5 v^4 e^{-\lambda v}}{4!}, \qquad v \geq 0.$$

$$\begin{aligned} \mathbf{P}(\{\text{Wallace's team wins}\}) &= \mathbf{P}(V > 32) = \int_{32}^{\infty} f_V(v)\,dv = \int_{32}^{\infty} \frac{\lambda^5 v^4 e^{-\lambda v}}{4!}\,dv \\ &= \left(\frac{1}{8}\right)^5 \frac{1}{4!} \int_{32}^{\infty} v^4 e^{-\frac{1}{8}v}\,dv \\ &= \frac{103}{3} e^{-4} \approx 0.629 \end{aligned}$$

where again the integral can be computed by repeated integration by parts.

Incidentally, the strategy proposed in part (d) is clearly not optimal for maximizing the probability of victory because the decision of whether or not to play Wallace with five fouls should depend also on the time remaining in the game. For example, assuming the present model, removing Wallace less than 32 minutes into the game does not make sense because it leads to certain defeat.

3. (a) The state diagram of the Markov chain is:



(b) State 5 is reachable from state 1 in a minimum of three transitions. Paths from state 1 to state 5 also include paths with a loop from 1 back to 1 (of length 3) and/or a loop from 5 back to 5 by way of state 7 (either length 2 or length 3). Therefore potential path lengths are $3 + 2m + 3n$, for $m, n \geq 0$. Therefore, $r_{15}(n) > 0$ for $n = 3$ or $n \geq 5$.

(c) From states 1, 2, and 3, all states are accessible because there is a non-zero probability path from these states by way of state 3 to any other state. From states, 4, 5, 6, and 7, paths only exist to states 5, 6, and 7.

(d) States 5-7 are recurrent because by the logic in (c), they can be reached from any other state. States 1-4 are transient; once the system has transitioned out of state 4, it cannot return to any state other than states 5, 6, or 7.

States 5, 6, and 7 form a recurrent class. Because it can be traversed from state 5 back to 5 in either 2 or 3 steps (as discussed in (b)), the system can return to state 5 after n steps for any $n \geq 2$; therefore it is aperiodic.

(e) One transition must be added to create a single recurrent class: for example, adding a transition from state 5 to state 1 would allow every state to be reached from every other state. Any transition from the recurrent class states 5,6, or 7 to any of the states 1, 2, or 3 would work.

4. (a) Given L_{n-1}, the history of the process (i.e., $L_{n-2}, L_{n-3}, \ldots$) is irrelevant for determining the probability distribution of L_n, the number of remaining unlocked doors at time n. Therefore, L_n is Markov. More precisely,

$$\mathbf{P}(L_n = j | L_{n-1} = i, L_{2-2} = k, \ldots, L_1 = q) = \mathbf{P}(L_n = j | L_{n-1} = i) = p_{ij}.$$

Clearly, at one step the number of unlocked doors can only decrease by one or stay constant. So, for $1 \leq i \leq d$, if $j = i-1$, then $p_{ij} = \mathbf{P}(\text{selecting an unlocked door on day } n+1 | L_n = i) = \frac{i}{d}$. For $0 \leq i \leq d$, if $j = i$, then $p_{ij} = \mathbf{P}(\text{selecting an locked door on day } n+1 | L_n = i) = \frac{d-i}{d}$. Otherwise, $p_{ij} = 0$. To summarize, for $0 \leq i, j \leq d$, we have the following:

$$p_{ij} = \begin{cases} \frac{d-i}{d} & j = i \\ \frac{i}{d} & j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

(b) The state with 0 unlocked doors is the only recurrent state. All other states are then transient, because from each, there is a positive probability of going to state 0, from which it is not possible to return.

(c) Note that once all the doors are locked, none will ever be unlocked again. So the state 0 is an absorbing state: there is a positive probability that the system will enter it, and once it does, it will remain there forever. Then, clearly, $\lim_{n\to\infty} r_{i0}(n) = 1$ and $\lim_{n\to\infty} r_{ij}(n) = 0$ for all $j \neq 0$ and all i, .

(d) Now, if I choose a locked door, the number of unlocked doors will increase by one the next day. Similarly, the number of unlocked doors will decrease by 1 if and only if I choose an unlocked door. Hence,

$$p_{ij} = \begin{cases} \frac{d-i}{d} & j = i+1 \\ \frac{i}{d} & j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Clearly, from each state one can go to any other state and return with positive probability, hence all the states in this Markov chain communicate and thus form one recurrent class. There are no transient states or absorbing states. Note however, that from an even-numbered state (states 0, 2, 4, etc) one can only go to an odd-numbered state in one step, and similarly all one-step transitions from odd-numbered states lead to even-numbered states. Since the states can be grouped into two groups such that all transitions from one lead to the other and vice versa, the chain is periodic with period 2. This will lead to $r_{ij}(n)$ oscillating and not converging as $n \to \infty$. For example, $r_{11}(n) = 0$ for all odd n, but positive for even n.

(e) In this case L_n is not a Markov process. To see this, note that $\mathbf{P}(L_n = i + 1 | L_{n-1} = i, L_{n-2} = i - 1) = 0$ since according to my strategy I do not unlock doors two days in a row. But clearly, $\mathbf{P}(L_n = i + 1 | L_{n-1} = i) > 0$ for $i < d$ since it is possible to go from a state of i unlocked doors to a state of $i + 1$ unlocked doors in general. Thus $\mathbf{P}(L_n = i + 1 | L_{n-1} = i, L_{n-2} = i - 1) \neq \mathbf{P}(L_n = i + 1 | L_{n-1} = i)$, which shows that L_n does not have the Markov property.