Exam Markov Decision Theory and Algorithmic Methods (153192)

January 28, 2008, 09:00-12:00 – LA 3520

This exam consists of 5 exercises with the following weights:

exercise	1	2	3	4	5
points	12	8	9	3	8

Motivate all your answers. When derivation is required, you most provide the derivation.

Good lucks

Exercise 1: Tandem queues with blocking

Consider a tandem network model of two servers. New customers arrive according to a Poisson process with rate λ and join an infinite waiting room in front of server 1. They receive service, exponentially distributed with rate μ_1, then they move to a second waiting room and eventually receive a second service, exponentially distributed with rate μ_2. The waiting room between the first and second server has finite capacity, and may contain (including the customer in service) at most C customers. If this waiting is full, server 1 stops service, and resumes when server 2 completes a service. We will model the system in matrix geometric form.

1. Give the transition diagram with transition rates of this model.
2. Specify the levels and phases of this model, and define the matrices B, A_0, A_1, A_2 of QBD process.
3. Give the stability condition in terms of the matrices A_0, A_1, A_2. The objective now is to write the stability condition explicitly in terms of λ, μ_1, μ_2 (Note that questions 3.1 and 3.2 are independent of the rest of questions):
   1. Find the equilibrium distribution of the irreducible Markov process of generator $A=A_0+A_1+A_2$.
   2. Derive the stability condition in terms λ, μ_1, μ_2, C.
4. Prove that **R**, QBD rate matrix, ***cannot*** be solved in closed form via the balance principle. Propose an iterative approach to solve numerically **R**.
5. Show that for the particular case where C=1 **R** ***can*** be solved in closed form via balance principle. Provide a matrix closed-form expression for **R** (Note: you do not need to compute the entries of **R** in closed form).
6. Specify the equilibrium distribution of level 0 denoted as Π_0 and give the expression for the equilibrium distribution.

Exercise 2: Population disease problem

Consider a population of N individuals. The disease is spread among the population when a contaminated individual enters in contact with another pure individual. All individuals can contact (meet) each others. The (random) time that separates two consecutive contacts between two individuals is called the inter-contact time. The distribution of inter-contact times of any two individuals is exponential with rate λ. For problem simplification, assume that contact durations between two persons are instantaneous and sufficient to transmit the disease. Furthermore, assume that all inter-contact times between the individual pairs are mutually independent. The time required for a contaminated person to recover from infection is distributed exponentially with rate μ except for exactly one individual that will always remain infected. Therefore, if there are n ($n \geq 1$) contaminated persons at time t then only $(n-1)$ persons can recover from the disease however the nth person remains contaminated (for ever). Note that a recovered person may be re-infected when contacting another infected person.

Let $N=\{N(t), t \geq 0\}$ record the number of infected individuals at time t. Without loss of generality assume that $N(0)=1$.

1. Show that N is an irreducible Markov process and give its transition diagram with transition rates. Give **Q** the generator of N.
2. Let $\mathbf{P} = I+\Delta\mathbf{Q}$, derive the conditions that Δ should satisfy such that **P** be a probability transition matrix of an irreducible Markov chain N*.
3. Prove that N and N* have the same equilibrium distribution. What does N* represent according to N?
4. Using **P**, what is the expected number of infected individuals at time t given that there is one infected person at time 0. Give comments on the numerical algorithm that computes this expectation (e.g. infinite sum truncation).

3. A decision maker observes a discrete-time system which moves between states s_1, s_2, s_3, and s_4 according to the following transition probability matrix:

$$P = \begin{bmatrix} 0.3 & 0.4 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.5 & 0.0 \\ 0.1 & 0.0 & 0.8 & 0.1 \\ 0.4 & 0.0 & 0.0 & 0.6 \end{bmatrix}.$$

At each point in time, the decision maker may leave the system and receive a reward of $R = 20$ units, or alternatively remain in the system and receive a reward of $r(s_i)$ units if the system occupies state s_i. If the decision maker decides to remain in the system, its state at the next decision epoch is determined by P. Assume a discount rate of 0.9 and that $r(s_i) = i$.

(a) Formulate this model as a Markov decision problem (MDP).

(b) Formulate the policy iteration algorithm for discounted MDPs.

(c) Use policy iteration to find a stationary policy which maximizes the expected total discounted reward.

(d) Find the smallest value of R so that it is optimal to leave the system in state s_2.

4. Consider a discounted Markov decison problem (MDP) with infinite horizon. Assume the expected discounted reward optimality criterion is used with discount factor λ. Let η and δ be two decision rules. Define a new decision rule ω by

$$\omega(s) = \begin{cases} \delta(s) & \text{if } v_\lambda^{\delta^\infty}(s) \geq v_\lambda^{\eta^\infty}(s), \\ \eta(s) & \text{if } v_\lambda^{\eta^\infty}(s) > v_\lambda^{\delta^\infty}(s). \end{cases}$$

Show that $v_\lambda^{\omega^\infty} \geq \max\{v_\lambda^{\delta^\infty}, v_\lambda^{\eta^\infty}\}$.

5. (a) Define a Markov reward process (MRP) and describe its relation to Markov decision processes (MDP) with the average expected reward optimality criterion.

(b) Define the gain g and bias h of the MRP.

(c) Show that if $g(j) = g(k)$ then $h(j) - h(k) = \lim_{N\to\infty}[v_N(j) - v_N(k)]$. Use this to explain why h is referred to as the relative value vector.

(d) Prove the following evaluation equations: If the MRP has transition matrix P and reward r then $(I - P)g = 0$ and $g + (I - P)h = r$.

(e) Consider an MDP with the average expected reward optimality criterion. Assume the sets A_s are finite. Then it is known that there exists a $g \in \mathbb{R}$ and an $h \in V$ for which

$$0 = \max_{d\in D^{\mathrm{MD}}} \{r_d - ge + (P_d - I)h\}.$$

Prove that if (g', h') is any other solution of the optimality equations then $g = g'$.