Fall 2017 Qualifier Exam:
Decision Science and Operations Research

September 18, 2017

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.
2. Indicate on the cover of *each* book the area of the exam (DSOR), your name, and the question answered in that book. On *one* of your books list the numbers of *all* the questions answered.
3. Return all answer books in the folder provided. Additional answer books are available if needed.
4. You are allowed *one* 8.5×11 sheet of paper with formulae.

SPECIFIC INSTRUCTIONS:

You must answer 5 out of 6 questions.

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced a problem has been stated incorrectly, mention this to the proctor. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

1.

Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ and $Q = \{x \in \mathbb{R}^n : Cx \leq d\}$ be two non-empty polyhedra.

(a) Write a linear programming formulation that solves the problem:

$$\min\{\|x - y\|_1 : x \in P, y \in Q\}$$

where $\|z\|_1 = \sum_{i=1}^n |z_i|$ is the 1-norm.

(b) Write the dual of the formulation you wrote in part (a).

(c) Justify that both the primal and dual problems have an optimal solution (you may use the strong duality theorem).

(d) Using the above primal/dual pair of linear programs, show that if $P \cap Q = \emptyset$, then there exists a vector $p \in \mathbb{R}^n$ such that $p^\top x < p^\top y$ for all $x \in P$ and $y \in Q$. [Hint: the vector p can be defined using an optimal dual solution.]

---

2. Suppose that women acquire human papillomavirus (HPV) (which can eventually lead to cervical cancer) according to a Markov chain. Most HPV infections resolve on their own, but some HPV infections lead to LSIL and HSIL, two preliminary disease stages before cancer is acquired. Each stage in the Markov chain represents one year. Women can also die from other causes (the Death state). The resulting Markov chain has six states with the following transition probability matrix:

P=

	Norm	HPV	LSIL	HSIL	Cancer	Death
Norm	0.94	0.05	0	0	0	0.01
HPV	0.5	0.39	0.09	0.01	0	0.01
LSIL	0.3	0.4	0.24	0.05	0	0.01
HSIL	0.01	0.05	0	0.92	0.01	0.01
Cancer	0	0	0	0	1	0
Death	0	0	0	0	0	1

You do NOT have to give numerical answers. Instead describe how you would solve for answers symbolically.

(a) Is this Markov chain ergodic? Explain your answer.

(b) Data suggests that 60% of patients start in the Norm state, 35% in the HPV state, and 5% in the LSIL state. What is the probability that a randomly selected patient will be in the HSIL state in (exactly) three years?

(c) Given that a patient is in the Norm state today, what is the probability that the patient will reach state the LSIL state <u>at least once</u> in the next three years?

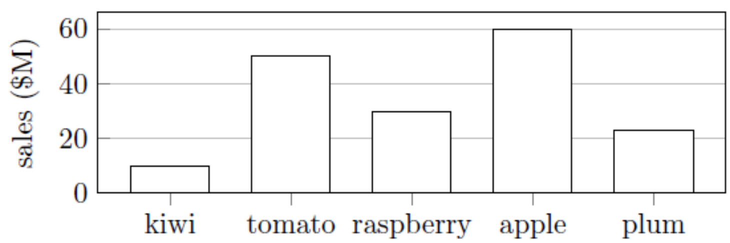
(d) Given that a patient starts in the Norm state, what is the expected time (in years) spent in the HPV state?

(e) Given that a patient starts in the Norm state, what is the probability that the HSIL state is ever reached?

3.

It is common knowledge that words/objects/entities have color associations. For example, *anger* is often associated with the color red. These associations are not one-to-one mappings, e.g. *strawberry* is also associated with the color red. The associations are not unique either; *apple* can be associated with red or green, and if we're talking about the company Apple Inc., the associations will be different still!

You are given a bar graph where each bar represents a different entity and your task is to choose colors to use for each of the bars. For example, the graph might look like the one below:



Your task is to choose colors for the bars in the graph so that each chosen color has a strong association with the category it represents. Suppose the labels for the bars in the graph are $\{b_1, \ldots, b_m\}$ and the colors at your disposal are $\{c_1, \ldots, c_n\}$. You have access to a dataset of color-category association strengths. The data is in the form of a table:

Category \ Color	c_1	c_2	$\cdots$	c_n
b_1	a_{11}	a_{12}	$\ldots$	a_{1n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
b_m	a_{m1}	a_{m2}	$\ldots$	a_{mn}

So a_{ij} is association strength between category b_i and color c_j. We'll assume all the data are normalized so that $0 \le a_{ij} \le 1$ and $\sum_j a_{ij} = 1$. In other words, you can think of the i^{th} row of the table as a distribution over colors for the category b_i. We'll assume $n \ge m$, so there are more colors than categories.

(a) Suppose we would like to assign the colors to the categories in a way that maximizes the total association strength of all pairs. For example, if we associate b_1 with c_6 and b_2 with c_1, then the total association strength is $a_{16} + a_{21}$. Note: you cannot assign the same color to two different categories. Formulate this optimization problem as a linear program that depends on the data $\{a_{ij}\}$. Be sure to explain why your model is correct and describe the variables, constraints, and objective function.

(b) The approach of minimizing total association strength doesn't work as well when several categories all have similar color association profiles. Instead, we'll look for a way to assign colors to categories such that the chosen pair has a high association strength and the non-chosen pairs have a low association strength. To this effect, define a new objective for all i, j:

$$h_{ij} = a_{ij} - \tau \max_{k \neq i} a_{kj}$$

Here, $\tau \geq 0$ is a parameter and the max is taken over all $k \in \{1, \ldots, i-1, i+1, \ldots, m\}$. The net effect of using such an objective is that b_i should be strongly associated with c_j and at the same time, c_j should not be strongly associated with any of the other b_k's for $k \neq i$. How should you modify your linear program to account for this new objective?

(c) Picking a different τ in the formula for h_{ij} generally leads to a different optimal assignment of colors to categories. Prove that when $m = 2$ and $n = 2$ (two categories and two colors), all values of $\tau \geq 0$ lead to the same solution.

4.

Given an undirected graph $G = (V, E)$ and a weight function $w : E \to \mathbb{R}$ on the edges, a *matching* $M \subseteq E$ is a subset of pairwise disjoint edges of G (i.e., every node of G is contained in at most one edge of M). The weight $w(M)$ of a matching M is defined as the sum of the weights of the edges in M, namely

$$w(M) = \sum_{e \in M} w_e.$$

In this setting the *maximum weight matching problem* asks to find a matching in the graph with maximum weight.

(a) Explain how the maximum weight matching problem can be solved in polynomial time if G is bipartite.

The *greedy algorithm* for the maximum weight matching problem proceeds as follows:

- Set $M := \emptyset$.
- Set $A := E$.
- While $A \neq \emptyset$, do:
  - Let e be an edge in A with highest weight.
  - Add e to M.
  - Remove from A all edges adjacent to e.
- Return M.

(b) Show an example in which the greedy algorithm does not find a matching with maximum weight.

(c) We now consider a restricted version of the maximum weight matching problem in which the weights of all edges are 1, hence a maximum matching M is simply a matching with maximum cardinality $|M|$. Notice that the greedy algorithm in this case chooses an arbitrary edge from A in every iteration.

Let OPT be the cardinality of the optimal solution and let M_g be the output of the greedy algorithm. Show that $\frac{|M_g|}{OPT} \geq 0.5$ (In other words, the matching that the greedy algorithm finds is at least half the size of an optimum one).

(d) For every $n \in \mathbb{Z}_+$ give a graph with at least n vertices for which $\frac{|M_g|}{OPT} = 0.5$.

---

5. Jobs arriving to a queueing system from a single customer, in a Poisson manner with rate λ. They are served with mean service time T. The system has a maximum capacity of four jobs—including both any job being served, and those jobs in the queue. Any jobs arriving when the system is full will just be rejected, and never enter the queue.

a) If service times are exponential, and only one job can be handled at a time, please draw a diagram showing the possible states of this system, and write down equations for the steady-state probabilities. (You do not need to solve those equations.)

b) The owner of the facility is not convinced that exponential service times are a good assumption, since there are fewer extremely short service times than would be predicted by the exponential model. Please describe another way that the service times could be modeled. In particular, is there a way of modeling the service times that would still leave the resulting system satisfying the Markovian property?

c) To simplify the process, the owner has decided to accept the exponential model after all. However, she is concerned that the average time in queue may be too long. Rather than accept all jobs and have many of them experience long wait times, the owner is considering turning away half the potential jobs, to keep average wait times short for those she accepts.

If she rejects every other job, is the resulting system still Markovian? What is the distribution for the time between accepted jobs in this case?

If she rejects half of all jobs at random by flipping a coin, is the resulting system still Markovian? What is the distribution for the time between accepted jobs in this case?

d) Another change that the owner is considering is having her customer submit jobs in batches of four. In this case, jobs would be accepted only when the system is empty. If batches of four jobs arrive in a Poisson manner, is the resulting system still Markovian? Is it reasonable to expect that jobs would arrive in a Poisson manner if the customer was asked to send jobs in batches?

---

6. A local package delivery company is exploring new policies for pricing package deliveries based on weight and volume of the package and the number of zones the package will travel. The table below shows the values of these parameters for 5 packages. When answering parts (b) and (c), be sure to show all your intermediate calculations, so that we can see the process you used, and also so that, in case you make a calculation error, we can identify it and still give the majority of credit if the process is correct.

Package	Weight (kg)	Volume (L)	Number of zones
1	8	6	1
2	7	10	2
3	10	12	1
4	15	16	2
5	5	7	3

(a) Use an appropriate plot to assess whether the Weight and Volume of packages can be modeled as statistically independent random variables. Draw the plot (label the axes) and make the best conclusion you can from the data.

(b) The current pricing policy is to charge $P_1 = 2 + \min\{W, V\} * N$ dollars, where W is the weight, V is the volume, and N is the number of zones. Use the data above to construct a 95% confidence interval for the expected value of a package delivery when pricing using policy P_1.

(c) The company is considering a new pricing policy , $P_2 = 1 + \max\{W, V\} * N$ dollars for a package of weight W, volume V and travelign N zones. Conduct an appropriate confidence interval (at the 95% confidence level) to determine how the expected value of the price using policy P_2 compares to the expected value of the price using P_1. Comment on whether your analysis suggests a (i) the statistically significant difference, and (ii) a practically significant differnece, between the two policies.