Fall 2018 Qualifier Exam:
Decision Science and Operations Research

September 17, 2018

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.

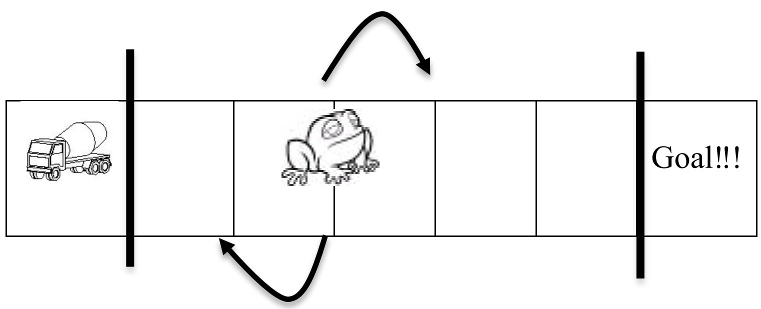
SPECIFIC INSTRUCTIONS:

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. A new video game called Frogette tracks the activity of a frog crossing a 5 lane highway toward her goal in a series of hops. Frogette moves one lane closer to her goal to the right with a probability of 0.7, otherwise she moves one lane to the left with a probability of 0.3. Each lane crossing is independent. The game ends when she either reaches the finish line or ends up at the start line (where she gets run over by a semi-truck. Poor Frogette).



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

(a) Formulate this as a Markov chain. What are the states? What is the transition probability matrix P?

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

(c) If she starts two lanes toward her goal (with 3 lanes to go), what is the probability she is three lanes toward her goal (one lane to the right in the picture above) in exactly three moves?

(d) If she starts two lanes toward her goal (with 3 lanes to go), what is the probability that she eventually reaches her goal on the right side of the highway?

(e) How many moves does she take, on average, before the game ends?

2. Assume that external virus attacks on a computer arrive according to a Poisson process with arrival rate λ. The probability that any particular virus attack will be successfully prevented by virus protection software is equal to P.

- Determine the probability that exactly k virus attacks will be successfully prevented in T time units.
- If we pick an interval of T time units, what is the probability that in that interval, there will be exactly R virus attacks that are successfully prevented and S virus attacks that are NOT prevented?
- Starting at time 0, a researcher studying computer security plans to observe the computer until at least one virus attack that is not prevented, and then a LATER attack that IS successfully prevented. Determine the expected amount of time that the researcher will have to observe the computer.
- Determine the probability mass function for the TOTAL number of virus attacks (regardless of whether they are successfully prevented) up to and including the third virus attack that is successfully prevented.
- Determine the expected value and probability mass function for the number of virus attacks that will be successfully prevented out of exactly N virus attacks in total.
- Estimate the arrival rate of virus attacks if the inter-arrival times between successive attacks (in days) are given as follows:

  10 days
  99 days
  15 days
  98 days
  20 days

A Simulation Question

Dumbledore has heard that you are a wizard of simulation, and he needs your help. Death Eaters are attacking Hogwarts, and he needs to build a simulation models of the attacks. As a first step, he needs to understand how often to expect an attack. Dumbledore enlisted Neville Longbottom to keep statistics about the number of attacks her hour. For the last 221 hours, Neville observed the frequency of attacks and collected the following data:

# Attacks	Observed Frequency
0 or 1	29
2 or 3	43
4 or 5	66
6 or 7	30
8 or 9	25
10 or 11	7

So 29 out of 200 times, either 0 or 1 attack occurred, 25 out of the last 200 hours, 8 or 9 attacks occurred, etc. Dumbledore believes that the Death-eaters attacks arrive according to a Poisson process. You will help him know if he can dispel this notion. Recall that the probability density function for a Poisson-distributed random variable is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Hermione Granger went to the restricted section of the Library, and retrieved 4 pages of tables that you may find useful when answering the following questions.

1. Explain what Dumbledore is assuming if the attack process is Poisson. Conversely, under what circumstances might a Poisson distribution for the number of attacks in an interval *not* be appropriate.

2. Estimate the average arrival rate parameter λ using the data given.

3. Employ an appropriate goodness of fit test to determine if one can reject the hypothesis that the arrival pattern is Poisson. Explain any estimates or approximations you are using, and what confidence level you have in your statement.

   In case the death-eaters have oblivated your memory, the Kolmogorov-Smirnov statistic is

   $$D = \max |F(x) - S_N(x)|,$$

   for the empirical distribution $S_N(x)$, and the and the chi-squared statistic is

   $$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

   where O_i and E_i are the observed and expected values in an interval.

4. **Linear Programming**

Recall the definitions of 1-norm and ∞-norm of a vector $x \in \mathbb{R}^n$:

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|, \qquad \|x\|_\infty := \max\{|x_i| : i = 1, \dots, n\}.$$

Consider the following optimization problem:

$$\begin{aligned} &\min\ c^\top x \\ &\text{s.t.}\ \ \|Ax + b\|_1 \le 1. \end{aligned} \tag{1}$$

In this formulation, the decision variables are $x \in \mathbb{R}^n$, and the given data consists of $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$.

(a) Formulate this problem as a linear program (LP) in inequality form and prove that your LP formulation is equivalent to problem (1).
Hint 1: You may use additional variables.
Hint 2: Recall that you can show that two maximization problems (A) and (B) are equivalent by showing: (i) For any feasible solution of (A) there is a feasible solution of (B) with objective value not lower, (ii) For any feasible solution of (B) there is a feasible solution of (A) with objective value not lower.

(b) Derive the dual LP, and show that it is equivalent to the problem

$$\begin{aligned} &\max\ b^\top z - \|z\|_\infty \\ &\text{s.t.}\ \ A^\top z + c = 0. \end{aligned} \tag{2}$$

What is the relation between the optimal value of z of problem (2) and the optimal value of the variables in the dual LP derived?

(c) Let x be feasible for (1) (i.e., $\|Ax + b\|_1 \le 1$) and let z be feasible for (2) (i.e., $A^\top z + c = 0$). Using only the weak duality theorem, what can you argue about the relation between $c^\top x$ and $b^\top z - \|z\|_\infty$?

5. **Optimization Modeling**

An *economy* consists of *sectors*. You can think of a sector as a process that consumes resources at the start of the year and produces other resources at the end of the year. We can also choose an *activity level* for each sector, which determines how much consumption and production happens in each sector.

Example. Here is an example with two sectors and three resources:

- Sectors: $\{\text{house-building}, \text{road-building}\}$
- Resources: $\{\text{wheat}, \text{brick}, \text{ore}\}$

At an activity level of 1, suppose we have the following:

- House-building consumes (1 brick, 1 ore) and produces (2 wheat, 2 brick, 2 ore).
- Road-building consumes (1 wheat, 1 brick) and produces (1 ore, 2 brick).

We can think of the consumption and production levels above as *rates*. To find the actual consumption and production, we multiply by the activity level. For example, if we choose an activity level of 100 for house-building and 50 for road-building in Year 1, then our economy behaves as follows:

- Total resources consumed by all sectors: (50 wheat, 150 brick, 100 ore)
- Total resources produced by all sectors: (200 wheat, 300 brick, 250 ore)

Every year, we must choose activity levels for the sectors, with the goal of making every sector *grow*. That is, we want the activity level of each sector to increase compared to the previous year. However, we cannot grow too fast: the consumption of a given year cannot exceed the production in the previous year. For example, if we decided to triple our activity levels for Year 2, this would require consuming (150 wheat, 450 brick, 300 ore). This is not possible because we produced an insufficient amount of brick and ore (300 and 250 respectively) in the previous year.

For this problem, we will look at optimizing growth over a two-year planning period. Specifically, we assume:

- There are m resources $i = 1, \ldots, m$ and n sectors $j = 1, \ldots, n$.
- For sector j, resource i is consumed at a rate b_{ij} and produced at a rate a_{ij}. These are fixed quantities known ahead of time.
- Sector j has activity level $x_j^{(1)}$ in Year 1 and $x_j^{(2)}$ in Year 2. These activity levels are things we must decide on.

- Year 1 activity levels are strictly positive and Year 2 activity levels are nonnegative.
- The consumption in Year 2 must not exceed the production in Year 1.

Define the *growth rate* of sector j as $x_j^{(2)}/x_j^{(1)}$. Our objective will be to maximize the *minimum growth rate*, which is the growth rate of the sector with the smallest growth rate. This ensures that every sector is growing. Finally, here are the problems:

(a) Formulate the above as an optimization problem. That is, specify the parameters, decision variables, constraints, and objective function.

(b) The formulation from Part (a) includes the strict inequalities $x_j^{(1)} > 0$ for $j = 1, \dots, n$. Explain why strict inequalities are generally undesirable in an optimization model. Also explain how and why the strict inequalities can be replaced by $x_j^{(1)} \geq 1$ without any loss of generality.

(c) Suppose we want to know whether it's possible to achieve a minimum growth rate of r. Explain how the problem of Part (a) can be reformulated as a linear program where r appears as a parameter.

(d) Consider maximizing the *total annual growth* $\left(\sum_{j=1}^n x_j^{(2)}\right)/\left(\sum_{j=1}^n x_j^{(1)}\right)$ instead. In this scenario, we allow activity levels in Year 1 to be zero, so long as the total activity in Year 1 is strictly positive. Formulate this problem as a linear program.

6. **Integer Optimization**

Let $S = \{v^1, \ldots, v^T\} \subseteq \mathbb{R}^n$ be a finite (possibly huge) set of points and consider the optimization problem:

$$z^* = \min\ c^\top x \tag{3}$$
$$\text{s.t. } Ax \geq b \tag{4}$$
$$x \in S \tag{5}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. Assume that this optimization problem has a feasible solution, so that z^* is finite. For $\lambda \in \mathbb{R}^m_+$, define:

$$z(\lambda) = \lambda^\top b + \min\ (c^\top - \lambda^\top A)x$$
$$\text{s.t. } x \in S$$

Finally, define:

$$z^{LD} = \max\{z(\lambda) : \lambda \in \mathbb{R}^m_+\}$$

The number of points that will be allocated to each part when grading are given in brackets at the beginning of each part.

(a) [2 pts] Show that $z(\lambda) \leq z^*$ for any $\lambda \in \mathbb{R}^m_+$.

(b) [4 pts] Recall that conv(S) is notation for the convex hull of S. Show that

$$z^{LD} = \min\ c^\top x$$
$$\text{s.t. } Ax \geq b$$
$$x \in \text{conv}(S).$$

[Hint: Start by formulating the problem $\max\{z(\lambda) : \lambda \in \mathbb{R}^m_+\}$ as a linear program, possibly by adding additional decision variable(s).]

(c) [2 pts] Provide an example where $z^{LD} < z^*$. [Hint: this can be done with $n = 1$ and a set S containing two points.]

(d) [1 pt] Suppose $S = \{x \in \mathbb{Z}^n_+ : Dx \geq d\}$. Use the result from part (b) to show that the $z^{LD} \geq z^{LP}$, where z^{LP} is the optimal value of the linear programming relaxation of (3)-(5):

$$z^{LP} = \min\ c^\top x$$
$$\text{s.t. } Ax \geq b$$
$$Dx \geq d$$
$$x \in \mathbb{R}^n_+$$

(e) [1 pt] Using the definition of S from part (d), suppose that the matrix D is totally unimodular and d is an integer vector. Again using results from previous parts, argue that in this case $z^{LD} = z^{LP}$.