Optimizing Healthcare Decision Making Lab

# Question 1

**Background**

1. Majority of cancer patients will go through radiation therapy (RT) at some stage of their treatment. RT uses high doses of radiation to kill cancer cells and shrink tumors. The following is a very simplified description of the process: (1) An initial CT scan of the patient is acquired; (2) the regions of interest (ROIs), which are the target and organs-at-risk (OARs), are defined on the CT; (3) a treatment plan is prepared, detailing the amount of dose to be delivered to each ROI; and (4) the radiation dose is delivered to the patient according to the treatment plan, which is often broken down into several sessions (referred to as fractionation of the treatment plan).
2. The quality of a treatment plan is determined through several measures, e.g., tumor control probability, which measures the likelihood of killing cancer cells, and normal tissue complication probability (NTCP) values, which measure the likelihood associated with certain side effects that may occur as a result of undesired irradiation of OARs. These measures are computed based on the positions of the ROIs determined within the initial CT scan.
3. At current, each patient is assigned only one treatment plan, which is prepared based on the initial CT scan. In reality, however, anatomical changes in patient’s body almost always occur during the treatment (especially, for highly fraction-

1

ated treatment plans). That is, after few days of treatment (and we do not know when), the ROIs start to move or shrink. Hence, it is wise to obtain more imaging throughout the dose delivery to observe the changes and adjust the treatment plan (i.e., adjust the dose to the new position of ROIs) if necessary. This is called replanning and can be as time- and resource-consuming as preparing the initial treatment plan.

**Problem Definition** Ideally, one would replan the treatment every day. However, this is impractical due to the resource limitation. Further, studies show that not everyone benefits from daily replanning. Hence, a clinician would like to find when is the right time (or when are the right times) to perform the replanning. This is a very critical question for highly fractionated treatment plans, e.g., 35-day treatment plans). The main issue is that the tumor and OAR’s response to the radiation *during the 35-day treatment* is stochastic (i.e., the movement and/or shrinkage). However, suppose you have access to daily imaging of the patient, which means you can observe, on a daily basis, the position of ROIs. How would you come up with a mathematical optimization model to aid the clinician? Suppose you aim at a treatment that minimizes the likelihood of a specific side effect.

# Question 2

An instance of bilevel optimization problem is defined as follows.

|  |  |
| --- | --- |
| min *f*(*x,y*) | (1a) |
| s.t *x* ∈X | (1b) |
| (*x,y*) ∈ *G*(*x,y*) | (1c) |

 *y* ∈ argmin {*g*(*x,y*) : *y* ∈Y(*x*)} (1d)

This model reflects a game, where a leader acts first by taking action *x*, while the follower acts next by taking action *y* with the full knowledge about the action taken by the leader. The leader minimizes an objective function *f* by varying *x*, while the follower minimizes another objective function *g*, given *x* as the action of the leader. The actions of both players are constrained. I propose the following approach: Replace the follower’s problem with the well-known Karush-Kuhn-Tucker system, and solve the resulting single-objective-function (nonlinear optimization) problem with off-the-shelf solvers. What can be claimed about the solution(s) obtained by this approach (1) when the follower problem is a linear program, and the leader problem is some generic mathematical program, and (2) when the follower problem is an integer linear program, and the leader problem is some generic mathematical program?

# Question 3

Assume there has been a clinical study on the effectiveness of a combination of two drugs, Drug 1 and Drug 2, to protect against a certain disease. Accordingly, you have access to a dataset with the following features. (Please see the attached data file to your email, and note that the data is subject to common *imperfectness* that one may encounter when working with real data.)

* **Drug 1 (mg)** (**Drug 2**) Total dosage of Drug 1 (Drug 2) given in miligrams.
* **Main Disease Status** Indicator of whether or not the main disease appeared at the end of the study.
* **Family Size**, **Age** (Self-explanatory)
* **Severity of the Relevant Disease** This indicates the severity of another disease, which is presumed to correlated with the main disease.
* **Surgery** This indicates whether or not the patient has gone through surgery. If so, it indicates the site of the surgery.
* **Smoking Status** (Self-explanatory)
* **Pack-year** This measures how many packs of cigarettes a year the patient has been smoking.
* **Diet Type** The patients are asked to follow one of five possible certain diets.
* **Gender**, **Race** (Self-explanatory)

Here is the question: you are asked to build a model that can compute the likelihood of occurring the disease in a patient as a function of the aforementioned features. Note that this question requires you to perform some coding; **I have no preference about your coding platform. If you add comments to your code, you help me understand the rationale of your model.**