IEOR 151: Service Operations Design and Analysis

Homework Assignment 1

(*Due September 23, 2019*)

Instructions:

1. Solve the questions by yourself and then self-grade your work.
2. Solutions of the questions are available on bCourses.
3. Submit the self-graded copy of your homework to bCourses by September 23, 2019.

1. Suppose the IT manager of a hotel chain would like to determine if the company should purchase a new electronic system to manage the sales/booking process. Purchasing the system will cost \$4600, and other hotel chains have found that using the electronic system leads to an average of 23 mistakes (e.g. lost sales opportunities) per day. If the current error rate is 27 mistakes per day, then purchasing the new system will have a net savings of \$5200 over the span of one year. The manager has decided to use a minimax hypothesis testing approach to answer this question. As a first step, you record the number of mistakes made over 20 days: 19, 23, 22, 37, 28, 18, 41, 36, 21, 27, 22, 42, 20, 33, 22, 10, 37, 12, 27, 33.

a. Assume that the number of mistakes per day is approximated by a Gaussian random variable with variance $\sigma^2 = 12$. Using a binary search and z-table, compute the threshold for this hypothesis test γ^* to within an accuracy of $\pm$ 0.1.
Hint: Use the following values for the minimax hypothesis test: $\mu_0 = 23$; $\mu_1 = 27$; $L(\mu_0; d_0) = 0$; $L(\mu_0; d_1) = \text{a} = 4600$; $L(\mu_1; d_0) = \text{b} = 5200$; $L(\mu_1; d_1) = 0$.

b. Should the manager purchase the new electronic system? Justify your answer.

Solution:

a. Consider the following comparison and recall that the goal is to select γ^* such that:

$$a(1 - \Phi(\frac{\sqrt{n}(\gamma^* - \mu_0)}{\sigma})) = b\Phi(\frac{\sqrt{n}(\gamma^* - \mu_1)}{\sigma})$$
$$4600(1 - \Phi(\frac{\sqrt{20}(\gamma^* - 23)}{\sqrt{12}})) = 5200\Phi(\frac{\sqrt{20}(\gamma^* - 27)}{\sqrt{12}})$$

Since $a < b$, binary search should be conducted on $[23, 25]$ and the best first guess is 24. Note, the required accuracy concerns γ rather than the difference between LHS and RHS. Using binary search, we obtain $\gamma^* = [24.94, 25]$. See the steps below.

Steps	γ	**LHS**	**RHS**
1	24.00	453.238	0.260
2	24.50	120.474	3.224
3	24.75	54.786	9.412
4	24.88	34.730	15.964
5	24.94	27.784	20.332

b. Given the decision rule:

$$\delta(X) = \left\{ \begin{array}{ll} d_0, & \text{if } \bar{X} \leq \gamma^* \\ d_1, & \text{if } \bar{X} > \gamma^* \end{array} \right\}$$

As $\bar{X} = 26.5 > \gamma^*$, the manager should choose d_1 and buy the new system.

2. *A portion of the following question is open-ended. Think thoroughly and provide good reasoning as we would like to understand your rationale.* Suppose you are the engineer in charged of building a spam filter for Gmail. The null hypothesis is that the incoming email is "ham" (the opposite of spam). Using Logistic Regression (a classification algorithm), you utilize the following decision rule to build the classifier:

$$\delta(X) = \left\{ \begin{array}{ll} d_0, & \text{if } g(X) \leq \gamma^* \\ d_1, & \text{if } g(X) > \gamma^* \end{array} \right\}$$

where g represents the model built via a training set.

a. Propose 3 predictors/statistics to collect in order to build a spam classifier.
Answer: 1-The portion of capitalized letters in the subject, 2-An indicator variable which denotes whether sender's email address is among "trusted" addresses, 3-The portion of special characters in the body of the email.

b. Propose a reasonable loss function. Justify your answer.
Answer: A 0-1 loss function or classification error.

c. Propose a metric to evaluate the degree of success achieved by the decision rule.
Answer: Number of false positives and false negatives.

d. *Bonus Question:* Graphically present the performance of such a classifier (as the threshold γ^* is increased) *Hint:* Analysts are often interested in the tradeoff between false positive rate and true positive rate.
Answer: Consider an Receiver Operating Characteristic (ROC) curve.

e. The product manager reviewed user statistics and found the number of type-I errors abnormally high. Propose the simplest method to correct the tendency.
Answer: We can simply adjust the threshold γ^* to make it harder to reject emails as spam in order to reduce false positives.

3. The supermarket that Mr. SpongeBob goes to is facing a slightly different problem. Every day, the supermarket will have to make a decision of how many apples to purchase from their supplier. Ideally, they want to purchase exactly the number of apples they will be able to sell the next day. However, demand for the next day is unknown. After some market research, they estimated that the demand of apple for tomorrow is normally distributed with a mean of 200 and a standard deviation of 60. Mr. SpongeBob purchases apples at \$0.60 per apple. The selling price is \$1 for each apple and unsold apples are sold at a discount price of \$0.20 per apple every evening. Some customers will choose to buy apples online if they come into the supermarket for apples only to find out that they don't have any left. This results in a goodwill cost of \$0.20 for each apple. What is the optimal number of apples that the supermarket should order in order to maximize total expected profit?

Solution:
Overage cost = c_o = 0.6 - 0.2 = \$0.4 per apple
Underage cost = c_u = 1 - 0.6 + 0.2 = \$0.6 per apple

$F(\gamma^*) = \dfrac{c_u}{c_u + c_o} = \dfrac{0.6}{0.6 + 0.4} = 0.60$. Hence, $z = 0.25$ and $\gamma = 200 + 0.25 \cdot 60 = 215$ apples.

4. Suppose we would like to select the optimal level of inventory of newspapers using a newsvendor model without production costs. Here, demand $X \sim U(100, 200)$ is in units of newspapers, holding cost is \$2.0 dollars per newspaper, and the sale price is \$1.5 dollars per newspaper. What is the optimal inventory level?

Solution:
The optimal inventory level γ^* is given by $F(\gamma^*) = \dfrac{1.5}{2.0 + 1.5} = 0.4286$.
First, observe that $\gamma^* \in (100, 200)$.
Hence, $F(\gamma^*) = \int_{100}^{\gamma^*} \dfrac{1}{200 - 100} du = 0.4286 \Rightarrow \gamma^* = 143$.