

## Optimization 2

### Project 1 – Stochastic Programming

#### Deliverables

One python file and one pdf document, submitted to Canvas. Your report should go into some detail about how you solved the problem, include some graphs that explain your results, and include relevant code chunks in the final output. 66% of your grade will be based on whether you get the problem right or not, the remaining 34% will be based on the quality of the presentation of your analysis. We will re-run your code with a new data set. If you don't get the right answer or your python file doesn't run, we will go through your code and give partial credit accordingly. The easier it is to read your code the easier it is for us to understand what you're doing, so use a lot of comments in your code!

#### Problem Overview

The newsvendor model that we solved in class is a relatively simplistic approximation of reality. The goal of this project is to extend the newsvendor model in a few ways to make a better approximation of reality.

In our first extension, we will assume that if you don't print enough newspapers to satisfy demand, then you can send a rush order to the printers to print all that you need. These rushed newspapers cost more to print, we will call the cost per rushed newspaper  $g$ , where  $g > c$ . Additionally, if you print more than the demand you must pay a disposal fee of  $t$  dollars per newspaper. You could alternatively think about this as receiving money for taking excess newspapers to a recycling center if  $t < 0$ , but for now we'll think of  $t > 0$ . Putting this together, the objective function is:

$$\max_q \frac{1}{n} \sum_{i=1}^n (pD_i - qc - g(D_i - q)^+ - t(q - D_i)^+)$$

where  $(x)^+ = \max(x, 0)$ . Part of this project will be to reformulate this as a linear program.

In our second extension of the NV model we will assume that price impacts demand linearly with error,  $D_i = \beta_0 + \beta_1 p + \epsilon_i$ , and jointly solve for the optimal price and quantity to print. This is in the model with disposal costs, and quick printing costs. To do this, you will be given a data set with price and demand data. You will fit a linear regression to this data, and you will assume the source of randomness is the residuals of this regression. This means once you find  $\beta_0$  and  $\beta_1$  you can find lots of demand data points for any one price: get all your residuals and plug them into the regression equation for that one price. That is, your data becomes just the residuals of your regression, instead of the original price/demand data. If you find  $\beta_0 = 1000$  and  $\beta_1 = -2$  and two of the residuals are 15 and -9, then for a price of \$2 you get 2 demand data points of  $1000 - 2 \cdot 2 + 15 = 1011$  and  $1000 - 2 \cdot 2 - 9 = 998$ .

You can see that when demand becomes a function of price like this you must multiply  $pD_i$  but this becomes  $p(\beta_0 + \beta_1 p + \epsilon_i)$ . This is then a quadratic function of  $p$ .

To solve these problems, you will have to create some dummy variables for each day,  $h_i$ , that represent some part of the objective; in class we let  $h_i$  represent the entire profit for each day. In this problem, you should set  $h_i$  to be the negative of cost on each day. Then the objective is the average of revenue plus negative cost, where the revenue part comes from  $pD_i$  and the negative cost comes from  $h_i$ . This way the quadratic part would stay in the objective instead of going down to a constraint when price impacts demand. You still need 2 constraints for the negative cost dummy variables, but these are linear constraints instead of quadratic. This is a little tricky to formulate because the problem is a

quadratic program, instead of a linear program. The reason we would define  $h_i$  to be negative cost is that we're maximizing over  $h$ , so we want negative costs to be as big as possible; we don't want big costs. Also, the constraints need to be upper bounds on  $h$ , not lower bounds. You can see this by drawing a similar graph to the one we drew in class when we did the original NV problem. Remember that costs are always positive in this model, so negative costs are always negative. Set your lb accordingly.

### Specifics

- 1) Data on price and demand are included in the assignment. Fit a linear regression model to this data set.
- 2) Let  $c=0.5$ ,  $g=0.75$ , and  $t=0.15$ . Using the residuals, assume the price is  $p=1$  and generate demand data.
- 3) Solve the optimal quantity to produce when  $p=1$ . This is not quadratic at all, just an LP.
- 4) Now let price impact demand and solve the resulting QP. What are the optimal price and quantity to print?
- 5) If you get in a bind and can't figure out how to formulate the QP you can formulate it as a general NLP, but that will result in a 10 point deduction from the project grade.
- 6) We are now interested to know how sensitive the optimal price and quantity are to our data set. Take a bootstrap sample of the original dataset. Go back and fit new beta's to the new bootstrapped dataset and redo step 4. Find the optimal price and quantity.
- 7) Repeat this process of getting new bootstrapped data, fitting new betas to the newly simulated data and finding the optimal price/quantity many times. Make histograms of the optimal price and quantity. Make a scatterplot with histograms on the x and y axis for this. You can do this quite easily using ggplot2. Make a histogram of expectation of profits.
- 8) You work for a publishing company. Your boss has been using the standard NV model to set printing decisions using the dataset given. Is your boss' model as good as this one? Could you increase revenue by switching models for production/pricing decisions? Describe the advantages and disadvantages of both techniques. Your boss is pretty technical and understands optimization, so don't be afraid to include quantitative material. Your boss is also busy, so be sure to include some visualizations to get the important points across. For the purpose of your report, you can assume that your boss is interested in the data posted with the project. Write this as if you were going to give it to your boss as a technical report that may be shared with the CEO. That means, make it look GOOD!
- 9) HINT: In order to help understand how to formulate the quadratic part of the problem, try doing this matrix multiplication:

$$(x_1 \quad x_2 \quad x_3) \begin{pmatrix} \beta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

What do you get? This just tells you how to do the quadratic part, there is still a linear part you'll need to figure out.