Estimation Theory Take Home Exam

Subhankar Chakraborty EE17B031

June 27 2020

Part I

Problem Summary

We are a part of an FMCG company that manufactures products to sell next month. We make five types of items. How much of an item we produce depends on the following factors:

- *Past history of sales*
- *Popularity of the item*
- *Perishability of the item*
- *Cost price versus selling price*

Simplifying assumptions made

The following simplifying assumptions have been made for the sake of convenience.

- The production of the five items is completely independent of each other. Workforce for each of the items specializes in that item itself. Lack of funds affects the allocation of funds to each of the items proportionately.
- Goods perish only during transportation. Once they reach a retail store, they do not perish for a month. At the end of the month, they become unusable and are discarded.
- There is no particular monthly or seasonal pattern in the distribution of sales of each item.
- Each of the five items follows a similar distribution, ie, the production model is the same for each one of them. The parameters of the model are different, however.

- The selling price is a constant for all purposes in a given market, ie, it is something we have to deal with and cannot decide ourselves in order to remain competitive against other competitors in the market. This is a reasonable assumption for all companies in a free market.
- The Cost price is assumed to be a **deterministic function** of the other separately mentioned factors like storage cost, shipping cost, ease of production, workforce needed, etc.

Approach

Figure 1 shows the flow chart for the overall process.

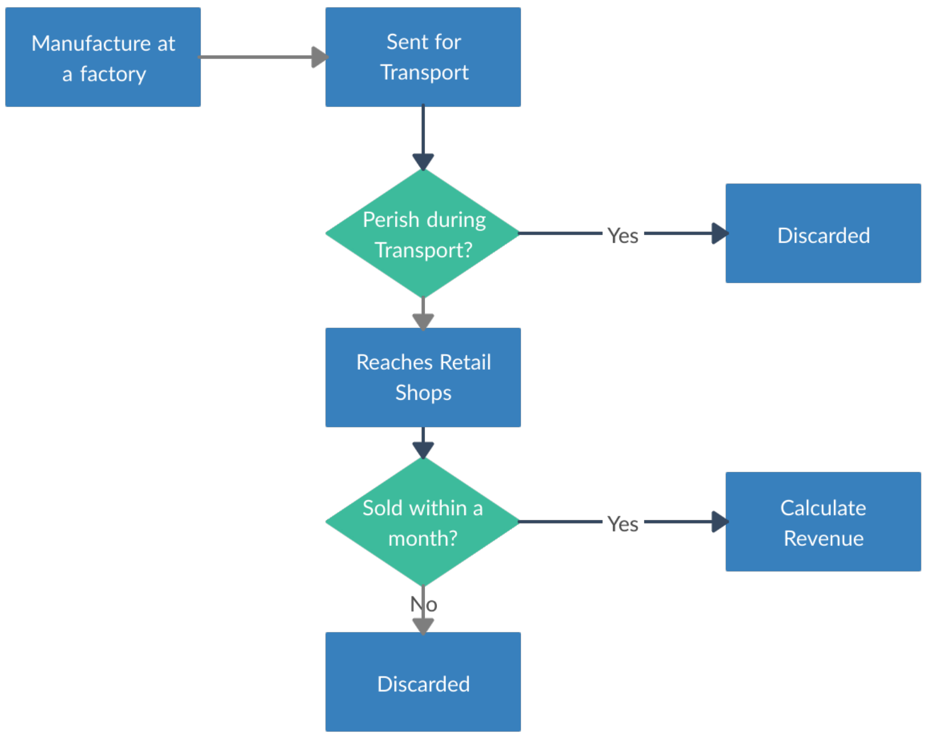

Figure 1: Process Flow Chart

Since we assumed the production and revenue of each item is completely independent of the other, **we show the mathematical model and the corresponding discussion for only one of the items for the sake of ease of**

notation and avoiding subscripts in every parameter. It can be scaled up to include all the items. The mathematical model based on the above flow diagram is defined below.

$$\mathbf{Revenue(R) = mS_p\text{-}nC_p\text{-}(K\text{-}m\text{-}n)C_p}$$
$$\mathbf{= mS_p\text{-}KC_p}$$
$$S_p = \text{Selling Price of the item}$$
$$C_p = \text{Cost Price of the item}$$
$$\text{K} = \text{number of items produced}$$
$$\text{n} = \text{number of items perished during transport}$$
$$\text{m} = \text{number of items sold within a month}$$

More on the Mathematical Model

We go through the above described mathematical model in detail here.

- The probability of an item perishing is given by $\mathbf{p_{perish}}$. Hence, the number of items that perish during transportation($\mathbf{n}$) is a binomial distribution with parameters K and p_{perish} , ie,

  $$\mathbf{n} \sim \mathbf{Binomial}(\mathbf{K}, \mathbf{p_{perish}})$$

  Some more details on p_{perish}:

  - We assume it to be a deterministic constant.
  - Its value can be calculated using historical data and Maximum Likelihood.

    $$\text{p}_{perish_{ML}} = \frac{\text{Total items perished during transport}}{\text{Total Items Produced}}$$

  - For the sake of convenience, we assume that items once purchased by consumers are immediately consumed and do not contribute to perished items.

- The probability of an item in the retail store being sold is assumed to have a prior beta distribution given by

  $$\mathbf{p_{sold}} \sim \mathbf{B}(\alpha, \beta)$$

  Some more details on p_{sold}:

  - α is set as the number of successes in the recent past, ie, the number of items which reached retail shops and were sold within the period of a month. This can be obtained using the recent popularity of the item.
  - β is set as the number of failures in the recent past, ie, the number of items that reached retail shops and were NOT sold within the period of a month. This too can be obtained using the recent popularity of the item.

  - For prediction, we use p_{sold} to be the mode of the prior distribution, ie, **unless mentioned otherwise,** $\mathbf{p_{sold}} = (\alpha - \mathbf{1})/(\alpha + \beta - \mathbf{2})$.

- Given $\mathbf{p_{sold}}$, **m** is assumed to be a binomial random variable with parameters K-n and p_{sold}, ie,

$$\mathbf{m} \sim \mathbf{Binomial}(\mathbf{K} - \mathbf{n}, \mathbf{p_{sold}})$$

Maximizing the Revenue

We can clearly see from our definition of the **R** that it is a function of **m**, which is a random variable. Hence, **R** is not a deterministic quantity and maximizing **R** is a vague sentence. We try and maximize $\mathbb{E}[\mathbf{R}]$, ie, the Expected value of **R**. $\mathbb{E}[\mathbf{R}]$ is given by

$$\mathbb{E}[\mathbf{R}] = \mathbb{E}[\mathbf{mS_p}\text{-}\mathbf{KC_p}]$$

The Expectation is calculated over the probability density of **m** and **n** based on our assumptions above. Given that $\mathbf{m} \sim \mathbf{Binomial}(\mathbf{K} - \mathbf{n}, \mathbf{p_{sold}})$ where **n** itself is a Binomial random variable, this expectation is hard to get in an analytic form. We can, however, use MC sampling to numerically calculate the value. Also, we must remember that the only value in our control here is **K** and to maximize $\mathbb{E}[\mathbf{R}]$, we need to essentially find the optimal value of **K**. A pseudo code for doing this is mentioned below. An implementation in Python is provided in the Appendix.

```
max_avg_revenue = 0
best_k = 0
for k=K_start; k<=K_end; k+=K_step:
    count = 0 #keeps track of the total size
    sum = 0 #for the Expectation
    N_array = samples_from_binomial_dist(k, p_perish)
    for n in N_array:
        M_array = samples_from_binomial_dist(k-n, p_sold)
        for m in M_array:
            sum = sum + m*Sp - K*Cp
            count = count + 1
        end for
    end for
    avg_revenue = sum/count
    if avg_revenue > max_avg_revenue:
        max_avg_revenue = avg_revenue
        best_k = k
    end if
end for
return best_k, max_avg_revenue
```

A careful look at the pseudo code will make it apparent that if the maximum expected revenue is never greater than zero, the optimal value of $\mathbf{K}$ returned is zero. This is intuitive as if we know in advance that we cannot make profit on an item, we will not produce any of it in the first place.

Part II

Problem Summary

Now suddenly there is a shutdown. The factory can only work with limited staff and for limited hours, the supply chain is disrupted, all workers have to be paid regardless of whether they are working full time or not, ships are not moving, truck movement is infrequent, people are buying only bare essentials. How will you change your manufacturing strategy? Is all your prior knowledge a complete waste or can some of it be still used?

Simplifying assumptions made

The following simplifying assumptions have been made for the sake of convenience.

- p_{perish} remains the same in this case as well. Note that based on our previous assumptions, p_{perish} was the probability of an item not surviving during transportation. This still does not change as it is an intrinsic property of the item being sold and is invisible to the demand in the market. How we account for the reduced demand in the market is that the items on reaching the retail shops have a lower probability of being sold than before within a period of a month, after which they become useless.

- The cost price per item definitely changes. This however can be determined from the changed labour and transportation rates, both of which we assume are available to us. The changed cost price per item is denoted as $\mathbf{C_{p1}}$.

- The selling price per item can also change as the market is not the same as before. However, as per our assumptions, we have no control over it and have to work with the market selling prices. The changed selling price per item is denoted as $\mathbf{S_{p1}}$.

- The overall model for the revenue is still valid.

Adjusting the model to a changed market

The model must reflect the fact that the demand is lesser after the imposition of lockdown. We account for it by changing the probability of an item being sold, ie, changing $\mathbf{p_{sold}}$. The other factors (except the cost price per item and the selling price per item) as explained above can remain the same.

Calculating the new probability of an item being sold

In this section we discuss how to calculate the new value of $\mathbf{p_{sold}}$ using the prior defined and recent data after the imposition of lockdown. We make the following assumptions.

- Since lockdown was imposed around March 17, we assume that we still have some unsold items by that time. Say the number is $\mathbf{q}$.
- Of the $\mathbf{q}$ items unsold, say we manage to sell $\mathbf{y}$ of them by the end of the month.
- The probability of an item being sold is independent to the number of items initially present.

From the points mentioned above, in our model this is a case of $\mathbf{m}$ being $\mathbf{y}$ given **K-n** is $\mathbf{q}$. We use the prior distribution on p_{sold} (denoted as $\mathbf{p}$ in this section for convenience) and the likelihood to calculate the posterior distribution of p_{sold}. As mentioned earlier, we update p_{sold} to be the **mode of the posterior distribution for future predictions.** The equations are shown below.

$$\mathbf{Pr}(p \mid m = y, K - n = q) \propto \mathbf{Pr}(m = y \mid p, K - n = q)\mathbf{Pr}(p \mid K - n = q)$$

Using the independence assumption declared above,

$$\mathbf{Pr}(p \mid m = y, K - n = q) \propto \mathbf{Pr}(m = y \mid p, K - n = q)\mathbf{Pr}(p)$$

Writing the above equation in the form of respective densities and ignoring the constants

$$\begin{aligned}\mathbf{Pr}(p \mid m = y, K - n = q) &\propto p^y(1-p)^{q-y}p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{y+\alpha-1}(1-p)^{q-y+\beta-1}\end{aligned}$$

The mode of this distribution ($\mathbf{p}^{`}$) is given by

$$\mathbf{p}^{`} = (y + \alpha - 1)/(q + \alpha + \beta - 2)$$

We set $\mathbf{p_{sold}}$ equal to $\mathbf{p}^{`}$ for future predictions. It is, however, important that $\mathbf{p_{sold}}$ be updated on a monthly basis. Given how rapidly the markets are changing as the country slowly comes out of the lockdown, the demand is expected to increase and companies must start producing more than they were during lockdown to increase their revenue.

Part III

What is the key take away from this problem and its solution?

Take aways

The following conclusions can be made.

- Real life problems often require completely different skills to be solved than the ones required to arrive at elegant mathematical proofs.
- Simplifying assumptions often hold the key to whether a problem can be solved or not. Real life problems have so many factors that it is impossible to account for everything.
- There is a tradeoff between fitting powerful mathematical models and keeping the model interpretable.
- Clever mathematical modeling along with simple tools like conjugate prior distributions and Bayes rule can be used to arrive at powerful and interpretable solutions.

References

1. A Bayesian Approach to Demand Estimation
2. Development of Bayesian production models for assessing the North Pacific swordfish population
3. Evaluating a Bayesian approach to Demand Forecasting with Simulation
4. Bayesian models of inductive learning
5. Part II: How to make a Bayesian model

Appendix

I : Python code to generate optimal K

A python code to generate optimal **K** is mentioned below. The source code can be found **here**.

```
def get_optimal_k(k_start=K_START, k_end=K_END,
p_perish=P_PERISH, p_sold=P_SOLD, step_size=STEP_SIZE,
size_of_array=SIZE_OF_ARRAY, cp=CP, sp=SP):
best_k = 0
max_avg_revenue = 0
for k in range(k_start, k_end, step_size):
    count = 0
    total_sum = 0
    N_array = np.random.binomial(k, p_perish, size_of_array)
    for n in N_array:
        M_array = np.random.binomial(k-n, p_sold, size_of_array)
        for m in M_array:
            revenue = m*sp-k*cp
            total_sum+=revenue
            count+=1
    avg_revenue = total_sum/count
    if avg_revenue>max_avg_revenue:
        max_avg_revenue = avg_revenue
        best_k = k
return max_avg_revenue, best_k
```