New York University Shanghai

Problem Set 5

BUSF-SHU 210: Business Analytics (Spring 2018)

Due at 10:00PM on Sunday, April 29
Solutions

Instructions

Please submit a zip file of a report together with the accompanying *R* codes on NYU Classes. The total score of the problem set is 7 points.

1. Duality in Linear Programming (1 point)

Consider the following LP:

$$(\mathcal{P})\ \min\ x_1 - 2x_2 + 3x_3$$
$$\text{subject to}$$
$$\begin{cases} x_1 \geq 1 \\ x_1 + 2x_2 + 4x_3 \geq 2 \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{cases}$$

Please briefly answer the following questions:

(a) (0.2 points) Does the LP ($\mathcal{P}$) have a finite solution? If yes, what is the optimal solution? What is the associated optimal objective function value?

All the codes for Question 1 can be found in the *R* script PS5_Q1.R.

We build the LP model in *R*. The solution reports an error code of 3, which means there is no finite solution for $\mathcal{P}$.

GRADING. 0.2 points in total. 0.1 points for building the LP model in *R*. 0.1 points for a correct answer. You will receive a full credit if you use a geometric or an algebraic argument.

(b) (0.3 points) What is the dual to the LP ($\mathcal{P}$)? Does the dual have a finite solution? If yes, what is the optimal solution to the dual? What is the associated optimal objective function value of the dual?

We first write the dual to $\mathcal{P}$:

$$
(\mathcal{D})\ \max y_1 + 2y_2
$$

subject to

$$
\begin{cases}
y_1 + y_2 \le 1 \\
0y_1 + 2y_2 \le -2 \\
0y_1 + 4y_2 \le 3 \\
y_1 \ge 0, y_2 \ge 0
\end{cases}
$$

Since the primal does not have a finite solution, the dual also does not have a finite solution.

GRADING. 0.3 points in total. 0.2 points for writing the dual. 0.1 point for figuring out that the dual also does not have a finite solution.

(c) (0.5 points) Consider the following Primal and Dual LPs:

$$
\text{(Primal)} \quad \max_{(x_1, x_2, \cdots, x_n)} C_1 x_1 + C_2 x_2 + \cdots C_n x_n
$$

subject to

$$
\begin{cases}
A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n & \le B_1 \\
A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n & \le B_2 \\
\cdots\cdots & \\
A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n & \le B_m \\
x_1 \ge 0, x_2 \ge 0, \cdots, x_n \ge 0 &
\end{cases}
$$

$$
\text{(Dual)} \quad \min_{(y_1, y_2, \cdots, y_m)} B_1 y_1 + B_2 y_2 + \cdots B_m y_m
$$

subject to

$$
\begin{cases}
A_{1,1}y_1 + A_{2,1}y_2 + \cdots + A_{m,1}y_m & \ge C_1 \\
A_{1,2}y_1 + A_{2,2}y_2 + \cdots + A_{m,2}y_m & \ge C_2 \\
\cdots\cdots & \\
A_{1,n}y_1 + A_{2,n}y_2 + \cdots + A_{m,n}y_m & \ge C_n \\
y_1 \ge 0, y_2 \ge 0, \cdots, y_m \ge 0 &
\end{cases}
$$

We define the vectors $C = (C_1, C_2, \cdots, C_n)'$ and $B = (B_1, B_2, \cdots, B_m)'$ ($'$ is the matrix transpose operator), and matrix

$$
A = \begin{pmatrix}
A_{11} & A_{12} & \cdots & A_{1n} \\
A_{21} & A_{22} & \cdots & A_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
A_{m1} & A_{m2} & \cdots & A_{mn}
\end{pmatrix}
$$

Suppose that the (Dual) has a finite optimal solution $y^* = (y_1^*, y_2^*, \cdots, y_m^*)'$. Show that, for any $x = (x_1, x_2, \cdots, x_n)'$ in the feasible set of (Primal) (i.e., $x \ge 0$ and $Ax \le B$), and any $y = (y_1, y_2, \cdots, y_m)'$ in the feasible set of (Dual) (i.e., $y \ge 0$ and $A'y \ge C$), we have

$$
C'x \le B'y^* \le B'y,
$$

i.e.,

$$\sum_{i=1}^{n} C_i x_i \leq \sum_{j=1}^{m} B_j y_j^* \leq \sum_{j=1}^{m} B_j y_j$$

Since the Dual has a finite optimal solution y^*, the primal also has a finite optimal solution x^*. Furthermore, $C'x^* = B'y^*$. Since x^* is the optimal solution to the primal, $C'x \leq C'x^*$ for all x in the feasible set of the primal. Since y^* is the optimal solution to the dual, $B'y \geq B'y^*$ for all y in the feasible set of the dual. Therefore, $C'x \leq C'x^* = B'y^* \leq B'y$ for all x in the feasible set of the primal and y in the feasible set of the dual.

GRADING. 0.5 points in total. 0.1 points for establishing the existence of the primal optimal solution x^*. 0.1 points for $C'x^* = B'y^*$. 0.1 points for $C'x \leq C'x^*$. 0.1 points for $B'y \geq B'y^*$. 0.1 points for putting everything together correctly.

2. Fundamental Theorem of Asset Pricing (2 points)

Assume that we have a bond, 2 Stocks (Stock 1 and Stock 2), a European call option written on Stock 1, and a European put option written on Stock 2. The strike price of the call option is $K_c = \$100$, i.e., the owner of the option has the right to exercise the option and buy the stock at the price \$100. The strike price of the put option is $K_p = \$120$, the owner of the option has the right to exercise the option and sell the stock at the price \$120. In period 0 (the current period), the price of the bond is \$100, the price of Stock 1 is \$100, and the price of Stock 2 is \$100. We assume that the current price of the call option is p_c whereas that of the put option is p_p. In period 1, there are 3 possible states: State 1 occurs with probability 0.3, State 2 occurs with probability 0.4, and State 3 occurs with probability 0.3. The cash flow of the bond is does not depend on which state we are in, and is equal to \$110. There are some volatilities in the stock prices of Period 1: The price of Stock 1 is \$125 in State 1, \$115 in State 2, and \$90 in State 3; The price of Stock 2 is \$150 in State 1, \$107 in State 2, and \$70 in State 3. The cash flow and current price information is summarized in the following table:

	Bond	Stock 1	Stock 2	Call Option	Put Option	Probability
State 1	\$110	\$125	\$150	C_{1c}	C_{1p}	0.3
State 2	\$110	\$115	\$107	C_{2c}	C_{2p}	0.4
State 3	\$110	\$90	\$70	C_{3c}	C_{3p}	0.3
Current Price	\$100	\$100	\$100	p_c	p_p	

Please briefly answer the following questions:

(a) (0.5 points) Formulate the investment strategy to search for arbitrage as a linear program. What is the dual for this linear program?

First, we calculate the cash-flows of each asset under different states. $C_{1c} = 25$, $C_{1p} = 0$, $C_{2c} = 15$, $C_{2p} = 13$, $C_{3c} = 0$, $C_{3p} = 50$. We definite the investment strategy as $(x_b, x_1, x_2, x_c, x_p)$, where x_b is the quantity for bond, x_i is the quantity for stock i, x_c is the quantity for the call option, and x_p is the quantity for the put option. Then, we can formulate the investment strategy in the following linear programming:

$(\mathcal{P})$ $\max[-100x_b - 100x_1 - 100x_2 - p_c x_c - p_p x_p]$ (current cash flow)

Subject to

$$110x_b + 125x_1 + 150x_2 + 25x_c + 0x_p \geq 0 \text{ (cash flow of State 1 is non-negative)}$$
$$110x_b + 115x_1 + 107x_2 + 15x_c + 13x_p \geq 0 \text{ (cash flow of State 2 is non-negative)}$$
$$110x_b + 90x_1 + 70x_2 + 0x_c + 50x_p \geq 0 \text{ (cash flow of State 3 is non-negative)}$$

Then, we can formulate the dual of this LP:

$$(\mathcal{D}) \ \min[0y_1 + 0y_2 + 0y_3]$$

Subject to

$$110y_1 + 110y_2 + 110y_3 = -100$$
$$125y_1 + 115y_2 + 90y_3 = -100$$
$$150y_1 + 107y_2 + 70y_3 = -100$$
$$25y_1 + 15y_2 + 0y_3 = -p_c$$
$$0y_1 + 13y_2 + 50y_3 = -p_p$$
$$y_1 \leq 0, y_2 \leq 0, y_3 \leq 0$$

Grading Scheme: 0.5 points in total. 0.3 points for formulating the primal LP. 0.2 points for writing the dual.

(b) (0.5 points) Determine the risk-neutral prices of the call and put options, (p_c^*, p_p^*), so that the market is arbitrage-free.

Solving the dual, we find that, if and only if, $y_1^* = -0.335$, $y_2^* = -0.258$, $y_3^* = -0.316$, $p_c^* = 12.25$, $p_p^* = 19.15$, the dual will have a finite solution.

Grading Scheme: 0.5 points in total. 0.4 points for solving the dual. 0.1 points for getting the correct prices.

(c) (0.5 points) Find the risk-neutral probability $P^* = (P_1^*, P_2^*, P_3^*)$. Does P^* depend on the real probabilities $(0.3, 0.4, 0.3)$? Suppose we have a new asset $\mathcal{A}$ whose cash-flow (of period 1) is C_1 in State 1, C_2 in State 2, and C_3 in State 3. What is the non-arbitrage price of $\mathcal{A}$ in period 0?

The risk-neutral probabilities P^* can be found using the formula $P_i^* = y_1^*/(y_1^*+y_2^*+y_3^*)$ $(i = 1, 2, 3)$. Hence, $P_1^* = 0.369$, $P_2^* = 0.284$, and $P_1^* = 0.348$. The risk-neutral probabilities are independent of the true probabilities $(0.3, 0.4, 0.3)$. The non-arbitrage price of $\mathcal{A}$ is $y_1^* C_1 + y_2^* C_2 + y_3^* C_3$ or $1/1.01 \times (P_1^* C_1 + P_2^* C_2 + P_3^* C_3)$.

Grading Scheme: 0.5 points in total. 0.2 points for obtaining the risk-neutral probabilities. 0.1 points for identifying that P^* are independent of the true probabilities. 0.2 points for pricing the new asset $\mathcal{A}$.

(d) (0.5 points) Suppose that $p_c = \$15$ and $p_p = \$15$. Try to find an arbitrage investment strategy.

The risk-neutral price for the call option is $p_c^* = 12.25 < p_c$. There is arbitrage opportunity. Suppose we short sell 1 unit of the call option (i.e., $x_c = -1$) and do not buy any put option (i.e., $x_p^* = 0$). Then, for an arbitrage strategy, we have the optimal objective value function of the following linear program is strictly positive.

$$\max[-100x_b - 100x_1 - 100x_2 + 15] \text{ (current cash flow)}$$

Subject to

$$110x_b + 125x_1 + 150x_2 \geq 25 \text{ (cash flow of State 1 is non-negative)}$$

$$110x_b + 115x_1 + 107x_2 \geq 15 \text{ (cash flow of State 2 is non-negative)}$$

$$110x_b + 90x_1 + 70x_2 \geq 0 \text{ (cash flow of State 3 is non-negative)}$$

It can be checked that $(x_b, x_1, x_2) = (40.94, -39.01, -14.18)$ is feasible for the above linear program and, it will result in a strictly positive cash flow of 2.25. This is an arbitrage strategy.

Grading Scheme: 0.5 points in total. 0.2 points for reformulating the linear program. 0.3 points for finding a correct arbitrage strategy.

Hint: An arbitrage investment strategy means a portfolio of assets that yield strictly positive cash flow in period 0 and non-negative cash flows for any future states.

3. Assigning Medical Sales Representatives (1 Point)

You are responsible for assigning 22 hospitals to 4 medical sales representatives (SRs). As the SRs need to visit the hospitals under their responsibility quite often, the goal of assignment is to minimize the total travel distance of all 4 SRs. We summarize the distance between each hospital and each SR's office in the following table (in KM). Note that SR 1 is based in Hospital 4, SR 2 is based in Hospital 14, SR 3 is based in Hospital 16, and SR 4 is based in Hospital 22.

Hospital Index	Distance to SR 1	Distance to SR 2	Distance to SR 3	Distance to SR 4	Workload
1	16.16	24.08	24.32	21.12	0.1609
2	19	26.47	27.24	17.33	0.1164
3	25.29	32.49	33.42	12.25	0.1026
4	0	7.93	8.31	36.12	0.1516
5	3.07	6.44	7.56	37.37	0.0939
6	1.22	7.51	8.19	36.29	0.132
7	2.8	10.31	10.95	33.5	0.0687
8	2.87	5.07	5.67	38.8	0.093
9	3.8	8.01	7.41	38.16	0.2116
10	12.35	4.52	4.35	48.27	0.2529
11	11.11	3.48	2.97	47.14	0.0868
12	21.99	22.02	24.07	39.86	0.0828
13	8.82	3.3	5.36	43.31	0.0975
14	7.93	0	2.07	43.75	0.8177
15	9.34	2.25	1.11	45.43	0.4115
16	8.31	2.07	0	44.43	0.3795
17	7.31	2.44	1.11	43.43	0.071
18	7.55	0.75	1.53	43.52	0.0427
19	11.13	18.41	19.26	25.4	0.1043
20	17.49	23.44	24.76	23.21	0.0997
21	11.03	18.93	19.28	25.43	0.1698
22	36.12	43.75	44.43	0	0.2531

Each hospital must be assigned to exactly one SR. Furthermore, the total workload of each SR should be between 0.8 and 1.2. The workload of each hospital is given in the last column of the above table. The workload of one SR is the sum of the workloads of all the hospitals assigned to him/her.

(a) (0.5 points) Formulate the SR assignment problem as an integer program model. What is the optimal SR assignment policy?

All the codes for Question 1 can be found in the R script PS5_Q3.R.

Define $x_{ij} \in \{0, 1\}$ as the decision variables, where $x_{ij} = 1$ refers to that SR i is assigned to Hospital j, and $x_{ij} = 0$ otherwise. The objective function is

$$\min \sum_{i=1}^{4} \sum_{j=1}^{22} d_{ij} x_{ij},$$

where d_{ij} is the distance between SR i and Hospital j, as shown in the Table.

There are several constraints for this integer programming problem:

- $\sum_{i=1}^{4} x_{ij} = 1$ $(j = 1, 2, \cdots, 22)$: Each hospital is assigned to exactly one SR.
- $0.8 \leq \sum_{j=1}^{22} w_j x_{ij} \leq 1.2$ $(i = 1, 2, \cdots, 4,\ w_j$ is the workload of hospital $j)$: The total work load of each SR i is between 0.8 and 1.2 (w_j is the workload of Hospital j).

We solve the model using R (see PS5_Q3.R). Solving the model, we find that SR1 is assigned to hospitals $\{4, 5, 6, 7, 8, 9, 12, 19, 20\}$, SR2 is assigned to hospitals $\{11, 13, 14, 18\}$, and SR3 is assigned to hospitals $\{10, 15, 16, 17\}$, whereas SR4 is assigned to hospitals $\{1, 2, 3, 21, 22\}$.

Grading Scheme: 0.5 points in total. 0.3 points for building an integer programming model. 0.2 points for getting the correct solution.

(b) (0.5 points) We would like to restrict the workload of each SR to the range between 0.9 and 1.1. Without resolving the model, can you discuss whether the total travel distance of all SRs will increase or decrease compared with the solution in part (a)? What is the new optimal SR assignment policy if the workload of each SR must be between 0.9 and 1.1?

If the workload for each SR is between 0.9 and 1.1, the constraints are more restrictive. Therefore, the optimal total travel distance will increase compared with the optimal solution in part (a).

We resolve the model using R (see PS5_Q3.R), and find that SR1 is assigned to hospitals $\{4, 5, 6, 7, 8, 9, 12, 19\}$, SR2 is assigned to hospitals $\{11, 13, 14, 18\}$, and SR3 is assigned to hospitals $\{10, 15, 16, 18\}$, whereas SR4 is assigned to hospitals $\{1, 2, 3, 20, 21, 22\}$.

Grading Scheme: 0.5 points in total. 0.3 points for discussing the changes of the optimal distance. 0.2 points for getting the correct new solution.

4. Verizon Pricing (2 points)

Verizon sells cell phone service, Internet access, and FIOS TV service to customers. Customers can buy any combination of these three products (or not buy any). The seven available product combinations include the following:

- Internet
- TV
- Mobile Phone
- Internet and TV
- Internet and Mobile Phone
- TV and Mobile Phone
- Internet, TV, and Mobile Phone

The data set `Verizon.csv` gives the amount 77 representative customers are willing to pay per month for each service. Each customer in this data set represents a segment of customers. Specifically, this data set has the following variables:

- *Customer*: The index of the representative customers, taking integer values from 1 to 77.
- *Internet*: The willingness-to-pay of the representative customer for the Internet service.
- *TV*: The willingness-to-pay of the representative customer for the TV service.
- *Mobile*: The willingness-to-pay of the representative customer for the mobile phone service.
- *Size*: The size of the segment the customer represents.

A customers surplus for a service/service bundle is

$$\text{Willingness-to-pay for the service/service bundle} - \text{Price of the service/service bundle}$$

In particular, a customer's willingness-to-pay for a service bundle is the sum of his/her willingness-to-pay for each individual service in the bundle. For each customer, we assume he/she will purchase the product that yields the maximum nonnegative surplus (ties are broken arbitrarily); if no product yields a nonnegative surplus the customer is assumed to purchase nothing.

Your goal is to design a pricing strategy for Verizon to maximize the total profit, assuming that the marginal cost of each service is 0.

(a) (1 point) Formulate the pricing problem of Verizon as an optimization model.

(b) (1 point) Use *R* to solve the optimization model you build in part (a). What is the optimal pricing strategy for the 7 services/service bundles listed above?

We define the following decision variables:

- $p_I \geq 0$: The price of Internet.
- $p_T \geq 0$: The price of TV service.
- $p_M \geq 0$: The price of the mobile phone service.
- $p_{IT} \geq 0$: The price of Internet-TV bundle.
- $p_{IM} \geq 0$: The price of Internet-Mobile phone bundle.
- $p_{TM} \geq 0$: The price of the TV-Mobile phone bundle.
- $p_{ITM} \geq 0$: The price of the Internet-TV-Mobile phone bundle.
- $r_{ij} \in [0, p_j]$: Payment of a customer in Segment i to purchase product j ($i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$).
- $x_{ij} \in \{0, 1\}$: The decision of customers in Segment i with respect to product j ($i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$). $x_{ij} = 1$ means a customer in Segment i purchases product j; otherwise $x_{ij} = 0$.
- $s_i \geq 0$: The surplus of customers in Segment i ($i = 1, 2, \cdots, 77$).

The objective function is the profit from selling Word, Excel, and the bundle:

$$\max \sum_{i=1}^{77} \sum_{j \in \{I,T,M,IT,IM,TM,ITM\}} n_i r_{ij},$$

where n_i is the size of segment i.

The following constraints should be satisfied under the optimal pricing strategy:

- The bundle price is lower than the sum of individual prices in the bundle:
  - $p_I + p_T - p_{IT} \geq 0$
  - $p_I + p_M - p_{IM} \geq 0$
  - $p_M + p_T - p_{TM} \geq 0$
  - $p_I + p_{TM} - p_{ITM} \geq 0$
  - $p_T + p_{IM} - p_{ITM} \geq 0$
  - $p_M + p_{IT} - p_{ITM} \geq 0$
  - $p_I + p_M + p_T - p_{ITM} \geq 0$
- $r_{ij} - p_j \leq 0$ ($i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$): The payment of a customer is at most the price of the product s/he purchases.
- $x_{iI} + x_{iT} + x_{iM} + x_{iIT} + x_{iIM} + x_{iTM} + x_{iITM} \leq 1$ ($i = 1, 2, \cdots, 77$).
- $s_i - R_{ij} + p_j \geq 0$ ($i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$). R_{ij} is the reservation price of a customer in segment i for product j.
- $r_{ij} \geq p_j - R_{\max}(1 - x_{ij})$ ($i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$). $R_{\max} = \max R_{ij}$.
- $s_i = \sum_{j \in \{I,T,M,IT,IM,TM,ITM\}} (R_{ij} x_{ij} - r_{ij})$ ($i = 1, 2, \cdots, 77$).
- $s_i \geq 0$, $p_j \geq 0$, $r_{ij} \geq 0$, $x_{ij} \in \{0, 1\}$ for all $i = 1, 2, \cdots, 77$ and $j = I, T, M, IT, IM, TM, ITM$.

Grading Scheme: 2 points in total. 0.2 points for the decision variables. 0.2 points for the objective function. 1.6 points for the constraints.

5. Pricing Strategy of Starbucks (1 point)

Starbucks has offered two coffee choices: Latte and Cappuccino. Assume Starbucks charges p_l for Latte and p_c for Cappuccino (we assume Starbucks only has one cup size). Starbucks estimated that the hourly demand for Latte would be

$$\frac{1500\exp(25-p_l)}{1+\exp(25-p_l)+\exp(35-p_c)}$$

And the hourly demand for Cappuccino would be

$$\frac{1500\exp(35-p_c)}{1+\exp(25-p_l)+\exp(35-p_c)}$$

This demand model is called the Multi-Nomial Logit (MNL) discrete choice model, which is very widely used to analyze customer behaviors and optimize pricing and assortment strategies.

Assume that the cost of making one Latte is 10 RMB and that of making one Cappuccino is 13 RMB. You goal is to help Starbucks optimize their pricing strategy so as to maximize the total hourly profit.

(a) (0.5 points) Formulate the pricing problem of Starbucks as a non-linear programming model. Solve the NLP model in R. What is the optimal pricing strategy (p_l^*, p_c^*) that maximizes the total hourly profit of Starbucks?

All the codes for Question 5 can be found in the R script PS5_Q5.R.

The decision variables are (p_l, p_c). The objective function is the hourly profit of Starbucks:

$$\pi(p_l,p_c)=(p_l-10)\times\frac{1500\exp(25-p_l)}{1+\exp(25-p_l)+\exp(35-p_c)}+(p_c-13)\times\frac{1500\exp(35-p_c)}{1+\exp(25-p_l)+\exp(35-p_c)}.$$

The constraints are that $p_l \geq 10$ and $p_c \geq 13$.

We solve the model using R (see PS5_Q5.R), and obtain that $p_l^* = 29.1$ and $p_c^* = 32.1$.

Grading Scheme: 0.5 points in total. 0.3 points for building an NLP model. 0.2 points for getting the correct solution using R.

(b) (0.5 points) For some marketing concerns, Starbucks do not want the price difference between Latte and Cappuccino to be larger than 2 RMB. What is the optimal pricing strategy $(\hat{p}_l^*, \hat{p}_c^*)$ that maximizes the total hourly profit of Starbucks under this constraint?

The decision variables and the objective function are the same as those in part (a). The constraints should be adjusted to that $p_l \geq 10$, $p_c \geq 13$, $p_l - p_c - 2 \leq 0$ and $p_c - p_l - 2 \leq 0$.

We solve the model using R (see PS5_Q5.R), and obtain that $\hat{p}_l^* = 30.1$ and $\hat{p}_c^* = 32.1$.

Grading Scheme: 0.5 points in total. 0.3 points for (re-)building the NLP model by adjusting the constraints. 0.2 points for getting the correct solution using R.