Appointment Scheduling Under Patient No-Shows by Simulation: A Simple Least-Squares Approach

Preprint · October 2017

DOI: 10.13140/RG.2.2.31123.12328

Jason Dou
University of Pittsburgh

Appointment Scheduling Under Patient No-Shows by Simulation: A Simple Least-Squares Approach

This version: April 12, 2017

1. Introduction

The health care industry represents approximately 15% of the gross domestic product of the United States(Gupta and Denton (2008)). Under rising demand and cost, the health-care system needs to make best use of its existing capacity (Robinson and Chen (2010)). Appointment scheduling system is used to improve efficiency of health care system. However there are usually significant number of possible scheduling arrangement which is hard to enumerate. In the literature, Jiang et al. (2015) applies an integer programming approach for Appointment Scheduling with Random No-shows and Service Duration. LaGanga and Lawrence (2007) use simulation analysis to develop and test scheduling rules. Zacharias and Pinedo (2014) develop several properties and theorems on how optimal policy for scheduling looks like. In this project, we focus on applying dynamic programming approach to develop optimal policy for appointment scheduling with no-show and overbooking.

2. Problem Formulation

We first assume there are M patients which are divided into D classes each day, the M_j patients of each class with the same no-show probability $p_j, j = 1, 2...D$. For the doctor, there are T time slots each day. A two dimensional matrix $\mathbf{x} = [x_{tj}]$ means there is x_{tj} class j patients scheduled at time slot $t, t \geq 1$ and $t \in N$. We want to have a schedule to minimize the cost of patients' waiting time, doctors' idle time, doctors' over time and rejection cost during booking process. By applying the modeling method in Robinson and Chen (2010), we define:

$$t_{\max} = \max\{t | x_t \geq 1\}, x_t = \sum_{j=1}^{D} x_{tj}$$

$$\pi_t(k) : \text{the probability that there will be } k \text{ patients at time slot } t$$

$$\bar{n} = \sum_{j=1}^{D}(1-p_j)M_j : \text{expected workload of the doctor per day}$$

$$z_t : \text{the maximum possible number of patients in the office during slot t}$$

Then we have:

$$z_t = (z_{t-1}-1)^+ + x_t, \text{ with } z_0 = 0$$

$$\pi_t(k) = p(k \text{ show up out of } x_t)\pi_{t-1}(0) + \sum_{j=(k+1-z_{t-1})^+}^{min(x_t,k)} p(j \text{ show up out of } x_t)\pi_{t-1}(k+1-j)$$
$$\text{for } k = 0,...,z_t, t = 1,...,T.\pi_0(0) = 1 \text{ and } \pi_0(k) = 1 \text{ for } k \geq 1$$

Now we take "Cancellation" and booking horizon into consideration. Suppose we have a finite discrete booking horizon N, and n goes from N to 0. We use state variable $S_n(\mathbf{x})$ to denote that at booking period N, the scheduling condition for the whole day schedule is $\mathbf{x}$. In each time period n, we assume that one and only one of the events occurs: (1) an arrival of a patient of class j; (2) a cancellation by a patient currently holding a reservation; (3) a null event. There is probability λ_{nj} a new patient of class j's scheduling request will come. There is probability $q_{tnj}(x_t)$ that a patient of class j cancels her appointment at time slot t.
So we have $\sum_{j=1}^{D}(\lambda_{nj} + \sum_{t=1}^{T} q_{tnj}(x_t)) \leq 1$, for all $\mathbf{x}$ and n. Then we define $A(\mathbf{x})$ as the total time cost given schedule $\mathbf{x}$, I as the idle time, O as the over time, W as the waiting time. We have:

$$I(\mathbf{0}) = T, W(\mathbf{0}) = 0, O(\mathbf{0}) = 0$$

$$I(\mathbf{x}) = t_{max} - 1 + \sum_{k=0}^{z_{t_{max}}} k\pi_{t_{max}}(k) - \bar{n}$$

$$O(\mathbf{x}) =$$
$$\begin{cases} 0, & \text{if } t_{max} \leq T - z_{t_{max}}, \\ \sum_{k=1}^{t_{max}+z_{t_{max}}-T-1}(k\pi_{t_{max}}(k+T+1-t_{max})), & \text{if } T - z_{t_{max}} + 1 \leq t_{max} \leq T \end{cases}$$

$$W(\mathbf{x}) = \sum_{t=1}^{t_{max}-1}\sum_{k=1}^{z_t}(k-1)\pi_t(k) + \sum_{k=1}^{z_{t_{max}}}(\sum_{j=1}^{k}(j-1))\pi_{t_{max}}(k)$$

We calculate the total expected cost as weighted sum of the above three costs:

$$A(\mathbf{x}) = I(\mathbf{x}) + \alpha W(\mathbf{x}) + \beta O(\mathbf{x})$$

We have:

$$A(\mathbf{0}) = T$$

Then we take rejection of appointment request into account. The decision we need to make is that in the time period n, if a patient comes for appointment, whether we should accept or reject her request. If we decide to accept, which time slot from 1 to T we should schedule for her. For rejection cost, we assume it is a non-increasing function r_n depends on n. That means the later at the booking stage, the easier for the clinic to reject appointment request. To make it simple, we assume r_n is a linear function goes from $r_N = 5\alpha$ to $r_0 = 0.1\alpha$. Estimating α and β(that is to say, understanding the relative importance of waiting cost, idle cost, overtime cost, and rejection cost) is a critical component for appointment scheduling, but we do not explore that for now Robinson and Chen (2011). We define $C_n(\mathbf{x})$ as the minimum expected cost of operating the system over period n to 1. Then they are determined recursively by:

$$C_n(\mathbf{x}) = \sum_{j=1}^{D} \lambda_{nj} \min[\min_t [C_{n-1}(\mathbf{x}+\mathbf{e}_{tj})], C_{n-1}(\mathbf{x}) + r_n]] + \sum_{t=1}^{T}\sum_{j=1}^{D} q_{tnj}(x_t) * C_{n-1}(\mathbf{x}-\mathbf{e}_{tj}) +$$

$$(1 - \sum_{j=1}^{D} \lambda_{nj} - \sum_{t=1}^{T}\sum_{j=1}^{D} q_{tnj}(x_t)) C_{n-1}(\mathbf{x})$$

For the initial states:

for all $\mathbf{x}$(under clinic's capacity):

$$C_1(\mathbf{x}) = A(\mathbf{x})$$

3. A Numerical Illustration of the Dynamic Programming Approach

As Samiedaluie et al. (2013) does for neurology ward of hospital, here we demonstrate the applicability of the above dynamic programming approach by developing a simulated data set to represent the patient scheduling flows in a relative small size. And as we see, as the number of patients increases, the number of doctor's schedule increase dramatically. The *curse of dimensionality* hinders us from getting optimal solution of the dynamic programming model. Then we propose an approximation method inspired by valuation of American options in Longstaff and Schwartz (2001) to find a good policy in large-scale instances of the problem, explained in detail in the following section.

3.1 When does Curse of Dimensionality matter?

There is well-known curse of dimentionality problem when making use of the Bellman's equation Powell (2009). In our problem, the state space of doctor's schedule $\mathbf{x}$ increases exponentially when number of time slots and patients increases.

1. Example of data scale in real hospitals' operations

2. How does the computation go beyond control when N and k grow in experiment.

4. Least Square Monte Carlo Method for Approximation

Koch (2017) mentions that in general, there are two classes of Approximate Dynamic Programming(ADP) for learning value function approximations: mathematical programming and simulation. And the literature on capacity control mainly focuses on the first class. As far as our knowledge, our work is the first to apply the least-squares based simulation approach to appointment scheduling problem. The work contributes to simulation-based ADP literature and also provide new perspective for appointment scheduling problem.

Longstaff and Schwartz (2001) propose a least squares Monte Carlo (LSM) approach for valuation of American options. The internal similarity(suitable expression?) between minimizing overall cost in appointment scheduling and optimal exercise of American options inspires us to apply the LSM approach to the optimization problem in appointment scheduling. Then from Longstaff and Schwartz (2001) we can represent the unknown functional form of C as a linear combination of a countable set of basis functions of the schedule $\mathbf{x}$. Basic functions can be from as simple as powers of the state variables, Fourier, trigonometric series, to more sophisticated ones like Laguerre, Hermite, Legendre, Chebyshev, Gegenbauer and Jacobi polynomials(Longstaff and Schwartz (2001)). Then the problem is to find the best coordinates of the basic functions. we are starting from using ordinary least squares, then weighted least squares, generalized least squares, and even generalized method of moments to estimate the coordinates. The algorithm is illustrated as following:

Algorithm 1 old version: Least Square Monte Carlo For Appointment Scheduling

procedure
  $C_0(\mathbf{x}) \leftarrow A(\mathbf{x})$ ▷ Initialization in backward induction
  for $n \leftarrow 1,2,3...N$ **do**
    if `a new appointment comes` **then**
      for $t \leftarrow 1,2,3...T$ **do**
        `Regress` $C_{n-1}(\mathbf{x}+\mathbf{e}_t)$ on $\mathbf{x}$
        `We get:`$C_{n-1}(\mathbf{x}+\mathbf{e}_t) = \alpha_{n,t}\mathbf{B}(\mathbf{x})$ ▷ $\mathbf{B}$ is the choice of basic functions
        $C_n(\mathbf{x}) = \min\{\alpha_{n,t}\mathbf{B}(\mathbf{x}), t = 1,2...T; C_{n-1}(\mathbf{x}) + r_n\}$ ▷ all the possible time slots for the new patient or reject the new patient
    else if `cancellation request of time slot` t `comes` **then**
        $C_n(\mathbf{x}) = C_{n-1}(\mathbf{x}-\mathbf{e}_t)$
    else `null event happens`
        $C_n(\mathbf{x}) = C_{n-1}(\mathbf{x})$

Algorithm 2 Least Square Monte Carlo For Appointment Scheduling

procedure INITIALIZATION: REGRESSION PARAMETERS SET UP
  use a random variable E following uniform distribution to realize appointment arrival each time slot
  use a ”initial” policy to make decisions(reject or not; which time slot to put the new appointment)
  get the pairs of cost C_n and appointment schedule $\mathbf{x}$.
  regress C_n on basic functions of appointment schedule $\mathbf{B}(\mathbf{x})$ by least square method, we get the parameters $\mathbf{a}_n$.
procedure MAIN ALGORITHM
  for new appointment arrival, do decision making based on $C_n(\mathbf{x}) = \min\{\mathbf{a}_n\mathbf{B}(\mathbf{x}), t = 1,2...T; C_{n-1}(\mathbf{x}) + r_n\}$ ▷ all the possible time slots for the new patient or reject the new patient

5. Estimate the Probability Distribution of Appointment Arrival and No-show

I do not understand how to use the Cox Proportional-Hazards Regression model to estimate the probability distribution of patient’s appointment arrival yet. Fox and Weisberg (2011)

References

Fox, J., S. Weisberg. 2011. Cox proportional-hazards regression for survival data in r. appendix to an r and s-plus companion to applied regression.

Gupta, D., B. Denton. 2008. Appointment scheduling in health care: Challenges and opportunities. *IIE transactions* **40**(9) 800–819.

Jiang, R., S. Shen, Y. Zhang. 2015. Integer programming approaches for appointment scheduling with random no-shows and service durations. *Social Science Electronic Publishing* .

Koch, S. 2017. Least squares approximate policy iteration for learning bid prices in choice-based revenue management. *Computers & Operations Research* **77** 240–253.

LaGanga, L. R., S. R. Lawrence. 2007. Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences* **38**(2) 251–276.

Longstaff, F. A., E. S. Schwartz. 2001. Valuing american options by simulation: a simple least-squares approach. *Review of Financial studies* **14**(1) 113–147.

Powell, W. B. 2009. What you should know about approximate dynamic programming. *Naval Research Logistics (NRL)* **56**(3) 239–249.

Robinson, L. W., R. R. Chen. 2010. A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing & Service Operations Management* **12**(2) 330–346.

Robinson, L. W., R. R. Chen. 2011. Estimating the implied value of the customer's waiting time. *Manufacturing & Service Operations Management* **13**(1) 53–57.

Samiedaluie, S., B. Kucukyazici, V. Verter, D. Zhang. 2013. Managing patient admissions in a neurology ward. Tech. rep., Tech. rep., McGill University, Desautels Faculty of Management, Montreal, Quebec.

Zacharias, C., M. Pinedo. 2014. Appointment scheduling with no-shows and overbooking. *Production and Operations Management* **23**(5) 788–801.

A. Experiment Report

We apply the algorithm into the following stylized example:

D = 2, number of classes of patients

p = (0.2,0.2), no-show probabilities for each class

T = 2, number of time slots in doctor's schedule

N = 3, number of booking time period

lmd(λ) = matrix(0.4, N, D), appointment arrival probability

q = array(0.01,dim=c(T, N, D)), appointment cancellation probability

r =(3,3), rejection cost for each class.

And the initial policy is to assign appointment to the first minimum slot, and never reject it. Given the above input, we get:

DPcost = 3.576 (the expected minimum cost from exact dynamic programming model)

LSMcost = 4 (the realized experimental cost from Least square simulation method).

In following experiment, there are five aspects I'd like to explore further:

1.Simulation rounds: now I only simulate 5 rounds, I will simulate more rounds.

2.Basic function choices: now I use powers of state variables, but as mentioned in section 4, there are more promising basic functions like Fourier, Laguerre...

3.Generate bigger test cases: implement automatic test case generation

4.Modify the initial policy according to Zacharias and Pinedo (2014).

5.Extend the simplified last period cost function