1 Appendix I: Estimation of Parameters and Uncertain Data

1.1 Number of Same-Day Requests

For estimating the number of same-day requests, we assume that arrival of requests follows a Non-homogeneous Poisson Process (NHPP) which suggests different mean values for Poisson process during different periods of time. The Poisson process is widely used for patient arrival in clinic scheduling studies [Cayirli and Veral, 2003]. NHPP is usually suggested in phone-call arrival process [Weinberg et al., 2007]. Since estimation of request is done before the scheduling decision, in practice, the clinic can use other methods for estimation. Here we assume that the Type 2 requests in block j is from a NHPP with intensity function in (25), where $\lambda_1, \lambda_2, \lambda_3$ are constants.

$$\lambda(j) = \begin{cases} \lambda_1 & \text{if } j \in [1, \lfloor m/4 \rfloor] \cup [\lfloor m/2 \rfloor + 1, \lfloor m/2 \rfloor + \lfloor m/4 \rfloor] \\ \lambda_2 & \text{if } j \in [\lfloor m/4 \rfloor + 1, \lfloor m/2 \rfloor] \\ \lambda_3 & \text{if } j \in [\lfloor m/2 \rfloor + \lfloor m/4 \rfloor + 1, m] \end{cases} \tag{1}$$

The number of requests received in block j denoted by s_j has a Poisson distribution with mean given by (2). The expected number of requests received in block j equals the mean of s_j.

$$E[s_j] = \int_{j-1}^{j} \lambda(j)\, \mathrm{d}j \tag{2}$$

We assume the arrival of Type 3 patients also follows NHPP with intensity function in (3) where ζ_1 and ζ_2 are constants.

$$\zeta(j) = \begin{cases} \zeta_1 & \text{if } j \in [1, \lfloor m/4 \rfloor] \cup [\lfloor m/2 \rfloor + 1, \lfloor m/2 \rfloor + \lfloor m/4 \rfloor] \\ \zeta_2 & \text{if } j \in [\lfloor m/4 \rfloor + 1, \lfloor m/2 \rfloor] \cup [\lfloor m/2 \rfloor + \lfloor m/4 \rfloor + 1, m] \end{cases} \tag{3}$$

In the same way, the expected number of walk-in patients that arrive in block j denoted by w_j is shown in (4).

$$E[w_j] = \int_{j-1}^{j} \zeta(j)\, \mathrm{d}j \tag{4}$$

1.2 Attendance

If we do not consider the punctuality and cancellation of patients, the attendance of patients can be assumed to follow Binomial distribution as suggested in literature [Muthuraman and Lawley, 2008, Cayirli and Veral, 2003]. Let p_1 be the no-show probability of one Type 1 patient, assume that the number of regular patients that arrive at the jth block follows Binomial Distribution $\nu_j \sim \mathrm{B}(r_j, 1-p_1)$ satisfying (5).

$$\Pr\{\nu_j \text{ patients arrive at block } j\} = \binom{r_j}{\nu_j}(1-p_1)^{\nu_j} p_1^{r_j-\nu_i}, \quad \nu_j = 1, \cdots, r_j \tag{5}$$

In the same way, if we assume each of the Type 2 patient has a no-show rate of p_2, then we have the distribution in (6).

$$\Pr\{\alpha_j \text{ patients arrive at block } j\} = \binom{a_j}{\alpha_j}(1-p_2)^{\alpha_j} p_2^{a_j-\alpha_j}, \quad \alpha_j = 1, \cdots, a_i \tag{6}$$

1.3 Number of Patients that Can be Served in a Block

In literature, the distributions suggested for service time include: Exponential, Uniform, Gamma, Weibull and a few others [Cayirli and Veral, 2003]. In this paper, we use the number of patients served in a block as a random variable. Based on this setting, we can derive the distribution of this random variable from the service time. Three distributions are considered here: Poisson, Discrete Uniform and distribution derived from exponential service time. For the first two distributions, the mean served number is $\frac{l_j}{\xi}$, where l_j is the length of block j. Let T_k^j be the sum of service time of the first k patients in block j, then we have the distribution in (7).

$$\Pr\{\tau \text{ patients are served in block } j\} = \Pr\{T_{\tau+1}^j > l_j \text{ AND } T_\tau^j \leq l_j\} \tag{7}$$

Assume that service time of each patient follows a homogeneous Exponential Distribution with mean ξ, so T_k^j follows a Gamma Distribution with shape parameter k and scale distribution ξ.

Then we can update the distribution as follows in (8).

$$\begin{aligned}\Pr\{\tau \text{ patients are served in block } j\} &= \Pr\{T^j_{\tau+1} > l\}\Pr\{T^j_\tau \le l\} \\ &= (1 - F(l; \tau+1, \xi))F(l; \tau, \xi)\end{aligned}, \quad \tau \ge 1 \tag{8}$$

where $F(*)$ is the *cdf* of Gamma Distribution. If we have unequal lengths of blocks, we can plug in different values for l_j accordingly.

2 Appendix II: Special Cases

2.1 Unequal Lengths of Blocks

If we assume that the blocks have unequal lengths, then estimation of number of patient requests received in a block should be adjusted with the length of the block. The distribution of number of patients that can be served in a block should also be updated. This situation has been discussed in detail in Section 1.3.

2.2 No-shows of Type 2 Patients

The SIP-i model is built based on the assumption of zero no-show rates of Type 2 patients. In reality, the same-day request patients can also have a certain rate of no-show [Peng et al., 2014]. In this situation, we can define a new random variable $\alpha := \{\alpha_j\}$ representing the number of Type 2 patients arriving with appointments. The distribution of α_j depends on the value of a_j which is a first-stage decision variable. Introducing this random variable brings endogenous uncertainty to this problem. It means that the decision variable will affect the distribution of the random variable. The definition of η_j should be changed into:

$$\eta_j := \nu_j + \alpha_j - \tau_j \tag{9}$$

With endogenous uncertainty, SIP-i can only be solved by stage-wise decomposition with special cuts which is discussed in another paper by the authors [Fu and Banerjee, 2020].

2.3 Cancellations of Same-day Request Patients

If we consider cancellations of patients, we need to consider additional assumptions and settings. Here we only consider cancellations of Type 2 and Type 3 patients. Cancellations of Type 1 patients will be discussed later. For the current day, given the index of current block as i, let (p, j, t) be a triple of index associated with every assigned patients so far. Suppose the decision of cancellation is made at the beginning of each block. For $t = 2, 3$, the triple denotes patients with type t that made the request at $(i-p)$th block of today and were assigned to the the $(i+j)$th block. Especially, cancellation for Type 3 patients means they leave the clinic before the assigned block. Let L be the length between the request received block and the block when the patient decides to cancel the appointment. Let α_{pjt} be the probability that patient with (p, j, t) who have not canceled the appointment by the current block will arrive for the appointment. So we have:

$$\alpha_{pj2} = \Pr\{L \geq i + j + 1, \text{arrive for appointment} \mid L \geq i\} \tag{10}$$

$$\alpha_{pj3} = \Pr\{L \geq i + j + 1 \mid L \geq i\} \tag{11}$$

Let a_{pj2} be the number of Type 2 patients who request in block $i - p$ and are assigned to block $(i + j)$. Let $\hat{a}_{pj2}$ be the number of Type 2 patient who have not canceled the appointment by the current block and arrive at the assigned block. b_{pj3} and $\hat{b}_{pj3}$ can be defined for Type 3 patients in a similar manner. We assume that $\hat{a}_{pjt}$ and $\hat{b}_{pjt}$ are random variables from binomial distribution as shown below:

$$\hat{a}_{pj2} \sim \mathrm{B}\big\{a_{pj2}, \alpha_{pj2}\big\} \tag{12}$$

$$\hat{a}_{0j2} \sim \mathrm{B}\Bigg\{\sum_{k=1}^{\hat{s}} x_{jk}, \alpha_{0j2}\Bigg\} \tag{13}$$

$$\hat{b}_{pj3} \sim \mathrm{B}\big\{b_{pj3}, \alpha_{pj3}\big\} \tag{14}$$

$$\hat{b}_{0j3} \sim \mathrm{B}\Bigg\{\sum_{t=1}^{\hat{w}} y_{jt}, \alpha_{0j3}\Bigg\} \tag{15}$$

Before solving SIP-i, we need to obtain the values of α_{pj2} and α_{pj3} through statistical inference. Then the original second-stage constraints should be changed from

$$q_{j+1} = \max\{\nu_j + q_j + a_j + b_j - \tau_j, 0\} \tag{16}$$

$$g_j = \max\{\tau_j - \nu_j - q_j - a_j - b_j, 0\} \tag{17}$$

into

$$q_{j+1} = \max\{\nu_j + q_j + \sum_p \hat{a}_{pj2} + \sum_p \hat{b}_{pj3} - \tau_j, 0\} \tag{18}$$

$$g_j = \max\{\tau_j - \nu_j - \sum_p \hat{a}_{pj2} - \sum_p \hat{b}_{pj3} - q_j, 0\} \tag{19}$$

Or we can combine the above two sets of constraints together to obtain:

$$q_{j+1} - q_j - g_j = \eta(\omega)_j \tag{20}$$

where $\eta(\omega)_j$ can be obtained by replacing α_j in (9) with $\sum_p \hat{a}_{pj2} + \sum_p \hat{b}_{pj3}$, i.e.

$$\eta_j := \nu_j + \sum_p \hat{a}_{pj2} + \sum_p \hat{b}_{pj3} - \tau_j \tag{21}$$

2.4 Cancellations of Type 1 Patients

Liu et al. [2010] study the cancellation under far-in-advance scheduling policy where the cancellation is assumed to be handled in a daily mode at the beginning of each day. In this paper, we need a new method to handle the combination of daily cancellation and block-wise cancellation. For cancellations of Type 1 patients, let p in the triple (p, j, t) denote the number of days since the day when the patient sent the appointment request until today. Let D be the length between the request received day and the day when the patient decides to cancel the appointment.

Let θ_{pj1} be the probability that the patient with $(p, j, 1)$ who has not canceled the appointment until the current block will make the appointment of today. Assume that the cancellation behavior of patients is independent of the appointment date. Then we have:

$$\theta_{pj1} = \Pr\{L \geq i+j+1, \text{arrive for appointment} \mid L \geq i, D \geq p\} \tag{22}$$

Let $\hat{\nu}_{pj1}$ be the number of Type 1 patients who make the visit in the assigned block, we have:

$$\hat{\nu}_{pj1} \sim \mathrm{B}\big\{r_{i+j}, \theta_{pj1}\big\} \tag{23}$$

Then in SIP-i model we replace ν_j in (21) with $\sum_p \hat{\nu}_{pj1}$, i.e.

$$q_{j+1} - q_j - g_j = \eta(\omega)_j \tag{24}$$

$$\eta_j := \sum_p \hat{\nu}_{pj1} + \sum_p \hat{a}_{pj2} + \sum_p \hat{b}_{pj3} - \tau_j \tag{25}$$

The information about ratio θ_{pj1} can be obtained through study of historical data.

2.5 Punctuality of Type 1 and Type 2 Patients

If we relax the assumption about patient punctuality, and assume that patients may arrive earlier or later than the assigned block, the SIP-i model can be modified accordingly. According to Cayirli and Veral [2003], in relevant literature, the punctual arrival of patients is processed in the following ways: (1) empirical distribution of arrival time, (2) exponential distribution of arrival time, (3) allow only one block earliness or lateness. Contrasting to these methods, here we assume that the clinic only tolerates one-block lateness but can handle all early arrivals. It implies three situations:

- If the patients arrive earlier, they will still be served in their original assigned blocks. Their waiting time before the assigned block will not incur any cost.

- If the patients arrive more than one block behind, then they will be treated as walk-in patients. Their original appointments will be taken as no-shows.

- If the patients arrive at the immediate successive block of their original assigned block, they will be served in the block when they arrive.

To handle lateness, let $\gamma := \{\gamma_j\}$ denote the number of patients who arrive at block j but was assigned to $j-1$. This uncertain data is associated with $r_{i+j-1} + a_{j-1}$. Let ι_1 be the probability that a patient will be late, ι_2 be the probability that a patient will be late by only one block. Then we have the probability that a patient can make the appointment at the assigned block as $1 - \iota_1$. If we also consider the cancellations analyzed above, we have the following updates in addition to the formulations in Section 2.3 and Section 2.4:

$$\hat{a}_{pj2} \sim \mathrm{B}\big\{a_{pj2}, \alpha_{pj2}(1-\iota_1)\big\} \tag{26}$$

$$\hat{a}_{0j2} \sim \mathrm{B}\Big\{\sum_{k\in K} x_{jk}, \alpha_{0j2}(1-\iota_1)\Big\} \tag{27}$$

$$\hat{b}_{0j3} \sim \mathrm{B}\Big\{\sum_{t\in T} y_{jt}, \alpha_{0j3}\Big\} \tag{28}$$

$$\hat{\nu}_{pj1} \sim \mathrm{B}\big\{r_{i+j}, \theta_{pj1}(1-\iota_1)\big\} \tag{29}$$

$$\gamma_j \sim \mathrm{B}\big\{r_{i+j-1} + a_{j-1}, \iota_2\big\} \tag{30}$$

In Step 2, we need to count the number of Type 1 and Type 2 patients that arrived at the current block but assigned to block $i-2$ or before, and add the number to estimated walk-in patient number. Then (25) should be replaced with:

$$\eta_j := \sum_p \hat{\nu}_{pj1} + \sum_p \hat{a}_{pj2} + \sum_p \hat{b}_{pj3} + \gamma_j - \tau_j \tag{31}$$

Value of ι_1 and ι_2 can be obtained through historical data.

References

T. Cayirli and E. Veral. Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12(4):519–549, 2003.

Y. Fu and A. Banerjee. Stochastic integer model based dynamic open-access clinic outpatients appointment scheduling under endogenous uncertainty. *working paper*, 2020.

N. Liu, S. Ziya, and V. G. Kulkarni. Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing and Services Operations Management*, 12:347–365, 2010.

K. Muthuraman and M. Lawley. A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions*, 40:820–837, 2008.

Y. Peng, X. Qu, and J. Shi. A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients. *Computers & Industrial Engineering*, 72:282–296, 2014.

J. Weinberg, L. D. Brown, and J. R. Stroud. Bayesian forecasting of an inhomogeneous poisson process with applications to call center data. *Journal of the American Statistical Association*, 102(480):185–1198, 2007.