Appointment Scheduling with Restricted People

1 Preliminary Study

The service time for customer i, ξ_i, stochastic with a mean of μ_i and a standard deviation of σ_i. The service times are mutually independent. For each customer $i = 1, \ldots, n$, we use A_i to denote the appointment time, $S_i = \max\{A_i, S_{i-1} + \xi_{i-1}\}$ denote the actual starting time of service. We assume that the customers will arrive at the appointed time. Especially, $A_1 = S_1 = 0$.

The waiting time for customer i is $S_i - A_i$, the total waiting time is $\sum_{i=2}^{n} \alpha_i (S_i - A_i)$, where α_i is the weight for customer i. The overtime is $(S_n + \xi_n - T)^+$ and the total idle time is $\sum_{i=1}^{n-1} [S_{i+1} - (S_i + \xi_i)] = S_n - \sum_{i=1}^{n-1} \xi_i$.

In the scenario with at least 2 customers overlapping in the waiting room, we can calculate the overlapping time. Let t_{ij} denote the overlapping time between two customers i and j. Then, $t_{i,j} = (S_i - A_j)^+$, indicating there are at least $(j - i + 1)$ customers waiting.

The duration when there are only $(j - i + 1)$ people from customer i to customer j are waiting is $t_{i,j} - t_{i,j+1}$, $i = 2, \ldots, n-1, j \geq i$.

Total overlapping time: $\sum_{i=2}^{n-1} \sum_{j=i}^{n-1} \gamma_{i,j} (t_{i,j} - t_{i,j+1})$

Problem to minimize the total time cost:

$$
\begin{aligned}
\min_{\mathbf{A}} \quad & E_\xi \left[\left(S_n - \sum_{i=1}^{n-1} \xi_i \right) + \sum_{i=2}^{n-1} \sum_{j=i}^{n-1} \gamma_{i,j} (t_{i,j} - t_{i,j+1}) + \beta (S_n + \xi_n - T)^+ \right] \\
\text{s.t.} \quad & S_i = \max\{A_i, S_{i-1} + \xi_{i-1}\} \\
& S_1 = 0
\end{aligned} \tag{1}
$$

To minimize the makespan:

$$
\begin{aligned}
\min_{\mathbf{A}} \quad & E_\xi (S_n + \xi_n) \\
\text{s.t.} \quad & E_\xi (t_{i,j} - t_{i,j+1}) \leq L_{ij}, i = 2, \ldots, n-1, j \geq i
\end{aligned} \tag{2}
$$

L_{ij} indicates constraint on the duration of $(j - i + 1)$ people waiting.

2 Model

We redefine the scheduling problem using the following notation. Let $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_n)$ denote the **appointment intervals**, where Δ_i is the time allocated between the start of customer i and customer

$i+1$. Let $\mathbf{A} = (A_1, \ldots, A_n)$ represent the **scheduled appointment times**, with $A_i = \sum_{k=1}^{i-1} \Delta_k$ (assuming $A_1 = 0$). Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be the **random service durations**, where Z_i is the stochastic service time for customer i.

The waiting time for customer i is recursively defined as:

$$\begin{aligned} W_i(\mathbf{Z}_{i-1}, \mathbf{\Delta}_{i-1}) &= [A_{i-1} + W_{i-1}(\mathbf{Z}_{i-2}, \mathbf{\Delta}_{i-2}) + Z_{i-1} - A_i]^+ \\ &= [W_{i-1}(\mathbf{Z}_{i-2}, \mathbf{\Delta}_{i-2}) + Z_{i-1} - \Delta_{i-1}]^+, \end{aligned}$$

where $[*]^+ = \max(*, 0)$. $W_1(\mathbf{Z}_0, \mathbf{\Delta}_0) = 0$.

Let W_{ij} denote the **simultaneous waiting duration** for customers i through j, meaning the length of time during which all customers from i to j are simultaneously waiting. This can be expressed recursively as: $W_{i,j}(\mathbf{Z}_{i-1}, \mathbf{\Delta}_{j-1}) = [W_{i-1}(\mathbf{Z}_{i-2}, \mathbf{\Delta}_{i-2}) + Z_{i-1} - \sum_{k=i-1}^{j-1} \Delta_k]^+$, where $W_i \equiv W_{i,i}$ is the individual waiting time for customer i. W_{ij} captures the time window where customers i to j all experience waiting simultaneously due to delays from earlier customers (1 to $i-1$) and insufficient buffer times.

The finish time for customer i is $T_i(\mathbf{Z}_i, \mathbf{\Delta}_{i-1}) = A_i + W_i(\mathbf{Z}_{i-1}, \mathbf{\Delta}_{i-1}) + Z_i$. We aim to minimize the total schedule span, i.e., T_n, subject to constraints on individual and group waiting times.

The formulation of the problem can be expressed as follows

$$\begin{aligned} \min \quad & E\left[T_n(\mathbf{Z}_n, \mathbf{\Delta}_{n-1})\right] \\ \text{s.t.} \quad & E\left[W_{i,j}(\mathbf{Z}_{i-1}, \mathbf{\Delta}_{j-1})\right] \le w_{ij}, i = 2, \ldots, n-1, j \ge i \end{aligned} \tag{3}$$

In this setting, w_{ij} is related with the number of customers from i to j, i.e., $j - i + 1$. We can use w_k to indicate the upper limit on the time when there are k customers waiting.

$T_n(\mathbf{Z}_n, \mathbf{\Delta}_{n-1}) = A_n + W_n(\mathbf{Z}_{n-1}, \mathbf{\Delta}_{n-1}) + Z_n$.

Lemma 1. *For any given realization of $\mathbf{Z}_n$, $T_n(\mathbf{Z}_n, \mathbf{\Delta}_{n-1})$ becomes shorter when some customer is scheduled to arrive earlier while the schedule for others remain unchanged.*

The optimal schedule can be obtained by minimizing Δ_i.

When $i = 1$, the first customer doesn't need to wait.

When $i = 2$, only one constraint $E\left[W_1(\mathbf{Z}_0, \mathbf{\Delta}_0) + Z_1 - \Delta_1\right]^+ \le w_1$ is applied, then $\mathbf{\Delta}_1^*$ can be obtained.

When $i = 3$, there are two constraints on the waiting time of the third customer.

$E\left[W_2(\mathbf{Z}_1, \mathbf{\Delta}_1^*) + Z_2 - \Delta_2\right]^+ \le w_1$.

$E\left[W_1(\mathbf{Z}_0, \mathbf{\Delta}_0) + Z_1 - \Delta_1^* - \Delta_2\right]^+ \le w_2$.

When $i = n$, there are $(n-1)$ constraints.

Proposition 1. *By solving the above problems sequentially, the optimal schedule can be obtained.*

Then we analyze these problems. The function on the left-hand side is decreasing in the variable Δ_i.

When $\Delta_1 = 0$, $E\left[W_1(\mathbf{Z}_0, \mathbf{\Delta}_0) + Z_1 - \Delta_1\right]^+ = E\left[Z_1\right]^+$. If $E\left[Z_1\right]^+ \leq w_1$, $\Delta_1^* = 0$; if $E\left[Z_1\right]^+ > w_1$, $E\left[Z_1 - \Delta_1^*\right]^+ = w_1$.

If Z_i follows from the exponential distribution with rate λ, $E\left[Z_1 - \Delta_1\right]^+ = \frac{1}{\lambda}e^{-\lambda\Delta_1}$, then

$$\Delta_1^* = \begin{cases} -\frac{\ln(\lambda w_1)}{\lambda}, & \text{if } \lambda w_1 < 1 \\ 0, & \text{if } \lambda w_1 \geq 1 \end{cases}$$

The optimal schedule for (3) is feasible for (4) because the feasible region of (3) is smaller than that of (4).

$$\begin{aligned} \min \quad & E\left[T_n(\mathbf{Z}_n, \mathbf{\Delta}_{n-1})\right] \\ \text{s.t.} \quad & E\left[W_{i,j}(\mathbf{Z}_{i-1}, \mathbf{\Delta}_{j-1}) - W_{i,j+1}(\mathbf{Z}_{i-1}, \mathbf{\Delta}_j)\right] \leq w_{ij}, i = 2, \ldots, n-1, j \geq i \end{aligned} \tag{4}$$

3 Some instances

We calculate and compare three situations, including waiting time constraint, overlapping time constraint, waiting time and overlapping time constraint.

4 Literature

1. Possible traits: heterogeneous customers, no-show, lateness, walk-in

Different models: objective: minimize the total cost, minimize the makespan (the departure time of the last customer).

Traditional Appointment Scheduling Model.

1. with overbooking and no-shows (partial punctuality)

- discrete n time slots.

- minimize the waiting cost, idle time and overtime costs.

- analyze three components seperately

2. Under a service-level constraint (waiting time threshold)

- makespan

- the optimal schedule can be obtained sequentially.

5 Deterministic Situation

Suppose there are n customers to be scheduled. If the hard constraint-that certain individuals cannot be in the waiting room simultaneously-cannot be satisfied, the number of customers to be scheduled must be reduced until a feasible schedule is achieved.

Each customer i has a service time $Z_i \in [\underline{Z}_i, \overline{Z}_i]$. The goal is to maximize the number of scheduled customers n. The appointment time of the last customer does not exceed a given deadline $\overline{T}$. The number of simultaneous waiting customers, $S(t)$, never exceeds the capacity N at any time t.

This yields the following deterministic optimization problem:

$$\begin{aligned} \max \quad & n \\ \text{s.t.} \quad & W_i(\mathbf{Z}_{i-1}, \boldsymbol{\Delta}_{i-1}) = [W_{i-1}(\mathbf{Z}_{i-2}, \boldsymbol{\Delta}_{i-2}) + Z_{i-1} - \Delta_{i-1}]^+ \\ & S(t) \leq N, \forall t \in [0, \overline{T}] \\ & A_n \leq \overline{T} \end{aligned} \tag{5}$$

To obtain the number of simultaneous waiting customers, we define $j^*(i)$ as the largest index j satisfying:

$$(\overline{Z}_{i-1} - \sum_{i-1}^{j-1} \Delta_k) > 0, i = 2, \ldots, n$$

where Δ_k represents the appointment interval for customer k. The number of simultaneous waiting customers can be obtained by $j^*(i) - i + 1$. Given a fixed schedule $\boldsymbol{\Delta}$, the maximum j^* can be computed explicitly.

When all customers are assigned their maximum service time ($Z_i = \overline{Z}_i$), the schedule that maximizes the number of customers n is obtained by solving:

$$n^* = \max \left\{ n \middle| \sum_{i=1}^{n} \overline{Z}_i \leq \overline{T} \right\} + N + 1,$$

where $\overline{T}$ is the deadline and N is the waiting room capacity.

Example 1. *$Z_i \in [20, 40]$ for each i. $\overline{T} = 120$. $N = 2$. $n^* = 3 + 2 + 1$. The optimal schedule is $A_1 = 0$, $A_2 = 40$, $A_3 = 80$, $A_4 = A_5 = A_6 = 120$.*

The counterpart (dual) problem aims to minimize the schedule time A_n of the last customer, given a fixed number of customers n, while respecting the waiting area capacity constraint. The optimization problem is formulated as:

$$\begin{aligned} \min \quad & A_n \\ \text{s.t.} \quad & W_i(\mathbf{Z}_{i-1}, \boldsymbol{\Delta}_{i-1}) = [W_{i-1}(\mathbf{Z}_{i-2}, \boldsymbol{\Delta}_{i-2}) + Z_{i-1} - \Delta_{i-1}]^+ \\ & S(t) \leq N, \forall t \in [0, A_n] \end{aligned} \tag{6}$$

Example 2. *$n = 6$. $Z_i \in [20, 40]$ for each i. $N = 2$. The optimal schedule is $A_1 = 0$, $A_2 = 40$, $A_3 = 80$, $A_4 = A_5 = A_6 = 120$.*

For any given realization of $\mathbf{Z}_n$, an optimal schedule is $A_1 = 0$, $A_i = \sum_{k=1}^{i-1} Z_k, 2 \leq i \leq n - N$, $A_{n-N+1} = \ldots = A_n = \overline{T}$.

6 Stochastic Situation

The soft constraint can be set as the expected number of simultaneous waiting people does not exceed certain number.

Or the probability of the largest number of simultaneous waiting people is less than a threshold.

$$\begin{aligned} \max \quad & n \\ \text{s.t.} \quad & E[S(t)] \leq N, \forall t \\ & A_n \leq \overline{T} \end{aligned} \tag{7}$$

$$\begin{aligned} \min \quad & E\left[T_n(\mathbf{Z}_n, \boldsymbol{\Delta}_{n-1})\right] \\ \text{s.t.} \quad & E\left[S(t)\right] \leq N, \forall t \end{aligned} \tag{8}$$

The constraint $E\left[S(t)\right] \leq N, \forall t$ is equivalent to $E\left[S(A(i))\right] \leq N, i = 2, \ldots, n$.

$S(A(i)) = \max\{j | W_{i-1} + Z_{i-1} - \sum_{i-1}^{j-1} \Delta_k > 0\} - i + 1$.

$T_i(\mathbf{Z}_i, \boldsymbol{\Delta}_{i-1}) = \sum_{k=1}^{i-1} \Delta_k + W_i(\mathbf{Z}_{i-1}, \boldsymbol{\Delta}_{i-1}) + Z_i$

$W_i(\mathbf{Z}_{i-1}, \boldsymbol{\Delta}_{i-1}) = [W_{i-1}(\mathbf{Z}_{i-2}, \boldsymbol{\Delta}_{i-2}) + Z_{i-1} - \Delta_{i-1}]^+$

If the constraint is hard, for each i, we have $\max\{j | W_{i-1} + Z_{i-1} - \sum_{i-1}^{j-1} \Delta_k > 0\} - i + 1 \leq N$. The constraints can be converted to $Z_{i-1} + W_{i-1} \leq \sum_{i-1}^{N-2+i} \Delta_k, i = 2 \ldots, n$.