Single-Server Appointment Scheduling in Healthcare

Shubo Sun

April 25, 2024

Appointment Scheduling in Hospitals

Why does appointment scheduling play an important role for hospitals?

- Optimal healthcare resource utilization
- Patient Satisfaction
- Revenue Management
- Streamlined Operations
- . . .

Problem Setup

We assume a number of N patients are scheduled for exactly *one* appointment within a single day.

- All patients are scheduled for the same doctor (i.e, this is a single-server queueing problem.)
- The queue follows a first-come-first-serve rule.
- The arrival times are $0 \leq t_1 \leq \cdots \leq t_N \leq 1$.
- The interarrival times x_i is then defined as $x_i = t_i - t_{i-1}$.
- We assume the service time follows exponential distribution with mean $1/\lambda$.

Performance Measures

How can we check if a schedule is good?

- Less waiting time for customers (patients)
- Less idle time for servers (doctors)

Problem Formulation

We define

- W_i: waiting time for the i-th patient.
- I_i: idle time for the i-th patient.
- B_i: service time for the i-th patient (which follows $\exp(\lambda)$).
- S_i: sojourn time (amount of time this patient is in the system) for the i-th patient. Then $S_i = W_i + B_i$.
- By **Lindley's recursion** [1],
- $I_i = \max\{t_i - t_{i-1} - W_{i-1} - B_{i-1}, 0\} = \max\{x_i - W_{i-1} - B_{i-1}, 0\}$;
- $W_i = \max\{W_{i-1} + B_{i-1} - (t_i - t_{i-1}), 0\} = \max\{W_{i-1} + B_{i-1} - x_i, 0\}$.
- $E(I_i) = t_i + E(W_i) - (t_{i-1} + E(S_{i-1})) = x_{i-1} + E(W_i) - E(W_{i-1}) - E(B_{i-1})$

Problem Formulation

Our goal is to minimize the total cost, which consists of the total amount of customers' waiting time and the server's idle time:

$$Z(x_1, \cdots, x_{N-1}) = c_w E\left(\sum_{i=1}^{N} W_i\right) + c_I E\left(\sum_{i=1}^{N} I_i\right). \tag{1}$$

where $c_w \geq 0$ and $c_I \geq 0$ are associated costs for waiting time and idle time. We assume $c_w + c_I = 1$. In addition,

Problem formulation

$$Z := \min_{x1,\cdots,x_{N-1}} c^I \sum_{i=1}^{N} (x_{i-1} + E(W_i) - (E(W_{i-1}) + E(B_{i-1}))) + c_w \sum_{i=1}^{N} E(W_i)$$

subject to $c_I + c_w = 1, E(B_1) = \cdots = E(B_N) = 1/\lambda,$

and $c_I, c_w >= 0, 0 \leq x_1 \leq \cdots \leq x_{N-1} \leq 1$

Proof of Convexity

- The service time B_i is independent of x_i. Therefore, we can remove $E(B_i)$ when minimizing Z (the objective function).
- Applying the Lindley's recursion on W_i, W_i is convex interms of the interarrival times $x_1, \cdots, x_{N-1}$.
- Therefore, Z is convex in terms of $x_1, \cdots, x_{N-1}$. This guarantees a unique minimum of Z.

Optimization: Calculation of $E(W_i)$

We apply the method in Pegden and Rosenshine [2] to compute $E(W_i)$:

- $E(W_i) = \sum_{j=1}^{i-1} \frac{j}{\lambda} P(N(t_i) = j)$.
- When $j > 0$, $P(N(t_i) = j) = \sum_{k=0}^{j+k-1} \frac{(\lambda x_{i-1})^k e^{-\lambda x_{i-1}}}{k!} P(N(t_{i-1}) = j + k - 1)$.
- When $j = 0$,

$$P(N(t_i) = j) = \sum_{k=1}^{i-1} P(N(t_{i-1})$$

$$= k - 1) \sum_{h=k}^{\infty} \left(1 - \sum_{h=0}^{k-1} \frac{(\lambda x_{i-1})^h e^{-\lambda x_{i-1}}}{h!} \right)$$

Therefore, $E(W_i)$ can be computed iteratively.

Optimization: Find the optimal x_i

Algorithm 2: Optimization based on coordinate descent

Output: Optimal $(x_1^*, \cdots, x_{N-1}^*)$
Initialization: pick $\mathbf{x}_{(0)} \in \mathcal{R}_+^{N-1}$. Set $c_I = \alpha \in [0, 1]$ and $B = 100$. Set $\epsilon = \infty$ and $t = 0$.
while $\epsilon >= 10^{-5}$ **do**
  for $i = 1, 2, \cdots, N-1$ **do**
    Compute the first and second order derivatives using Newton's method.
    Update $x_i^{(t)}$ using Equation 6
  end for
  set $\epsilon = |\hat{Z}_{(t)} - \hat{Z}_{(t-1)}|$
end while
Return optimal $\mathbf{x}^*$ and Z^*

We note that this also works for the patient-non-punctuality case. Due to the randomness of $\mathbf{x}(0)$, we offer $\hat{x}^* = \frac{1}{B_s} \sum_{s=1}^{B_s} \mathbf{x}_s^*$ as an estimation of $\mathbf{x}^*$.

Scheduling for 10 patients

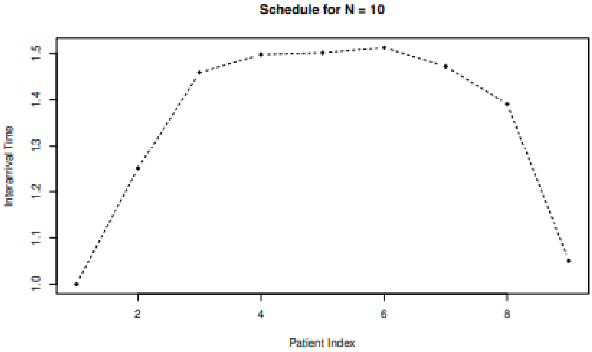


Fig. 1: Interarrival Times for $N = 10$

Scheduling for 15 patients

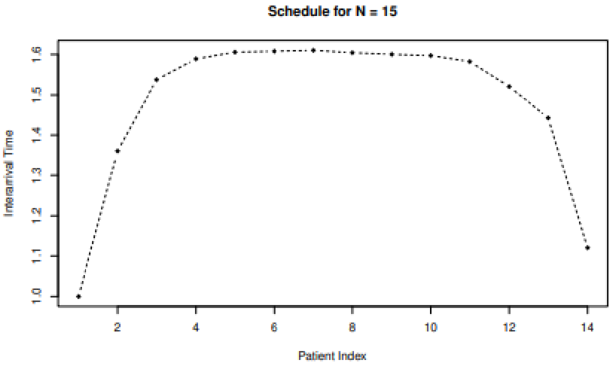
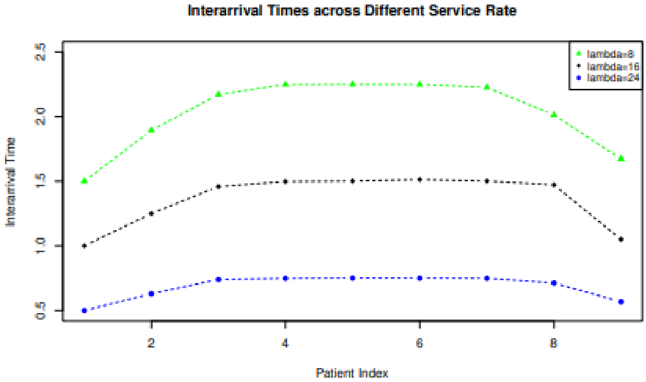
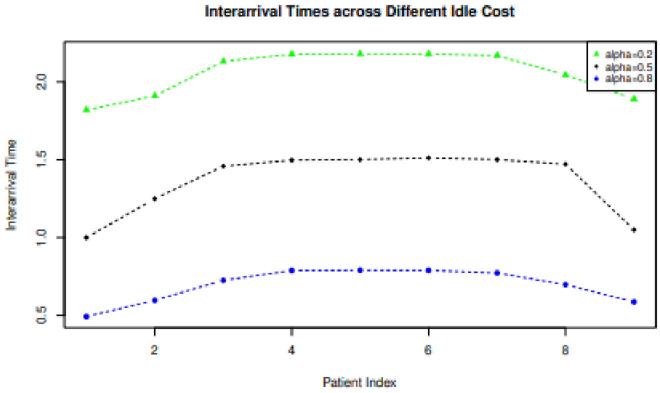


Fig. 2: Interarrival Times for $N = 15$

Interarrival Times for different Service Rate



Interarrival Times for different Idle Time Cost



References

- [1] Lindley, D. V. (1952). "The theory of queues with a single server". Mathematical Proceedings of the Cambridge Philosophical Society. 48(2):277–289.
- [2] C.D. Pegden, M. Rosenshine. Scheduling arrivals to queues. Computers Ops Res., 17 (1990), pp. 343-348.