A Flexible and Optimal Approach for Appointment Scheduling in Healthcare

APPENDIX: RECURSIVE PROCEDURE

Phase-type distributions are characterized by an integer $m \in \mathbb{N}$, an m-dimensional vector $\boldsymbol{\alpha}$ whose entries sum to 1, and an $(m \times m)$-dimensional transition rate matrix $\boldsymbol{S}$ whose entries satisfy $s_{ii} < 0$, $s_{ij} \geqslant 0$ and $\sum_{j=1}^{m} s_{ij} \leqslant 0$. Informally, the phase-type random variable is then the time it takes for a Markov process with rate matrix $\boldsymbol{S}$ to reach an absorbing state, starting in the initial distribution $\boldsymbol{\alpha}$. As can be checked easily, in the phase-type fit for ϱ smaller than 1 (with a mixture of Erlangs), the parametrization is $m = K$, $\alpha_1 = 1$, $\alpha_2 = \ldots = \alpha_K = 0$, whereas all entries of $\boldsymbol{S}$ equal 0, except $s_{ii} = -\mu$ for $i = 1, \ldots, K$, $s_{i,i+1} = \mu$ for $i = 1, \ldots, K-2$, and $S_{K-1,K} = (1-p)\mu$. In the case that ϱ is larger than 1, in which the phase-type fit is done by a hyperexponential random variable, we have to choose $m = 2$, $\alpha_1 = 1 - \alpha_2 = p$, $s_{12} = s_{21} = 0$, and $s_{ii} = -\mu_i$ for $i = 1, 2$.

Define the bivariate process $\{N_i(t), K_i(t), t \geq 0\}$ for patient $i = 1, \ldots, n$, in which $N_i(t) \in \{0, \ldots, i-1\}$ records the number of patients in front of the i-th patient t time units after her arrival (i.e., the queue length), and $K_i(t) \in \{1, \ldots, m\}$ represents the phase of the patient in service t time units after her arrival (which could be considered as the 'server status'). For each patient i at $t \geq 0$, where $j = 0, \ldots, i-1$, and $k = 1, \ldots, m$, we define the probabilities $p_{j,k}^{(i)}(t) = \mathbb{P}\left(N_i(t) = j, K_i(t) = k\right)$. Our objective is to evaluate the following vector of dimension mi:

$$\boldsymbol{P}_i(t) := \left(\left(p_{i-1,1}^{(i)}(t), \ldots, p_{i-1,m}^{(i)}(t)\right), \ldots, \left(p_{0,1}^{(i)}(t), \ldots, p_{0,m}^{(i)}(t)\right)\right).$$

In other words, the mi-dimensional vector $\boldsymbol{P}_i(t)$ keeps track of the probabilities corresponding to the queue and the phase of the patient in service t time units after patient i has entered. Let $p_{0,0}^{(i)}(t)$ denote the probability that patient i has left the system t time units after her arrival. The sojourn-time distribution of the i-th patient thus follows by noting that

$$F_i(t) := \mathbb{P}(S_i \leq t) = p_{0,0}^{(i)}(t) = 1 - \sum_{j=0}^{i-1}\sum_{k=1}^{m} p_{j,k}^{(i)}(t) = 1 - \boldsymbol{P}_i(t)\mathbf{1}_{mi},$$

in which $\mathbf{1}_{mi}$ is an all-one vector of dimension mi. Recall that all objective functions can be evaluated when knowing the distribution of the sojourn times $S_1, \ldots, S_n$; to this end, realize that, for $k_1, k_2 \in \{1, 2\}$,

$$\mathbb{E}\left[I_i^{k_1}\right] = \mathbb{E}\left[(x_{i-1} - S_{i-1})^{k_1}\mathbb{1}\{S_{i-1} < x_{i-1}\}\right],$$
$$\mathbb{E}\left[W_i^{k_2}\right] = \mathbb{E}\left[(S_{i-1} - x_{i-1})^{k_2}\mathbb{1}\{S_{i-1} > x_{i-1}\}\right].$$

The iterative procedure to evaluate $\boldsymbol{P}_i(t)$ is best explained by describing it on a patient-by-patient basis.

- For the first patient, arriving at time zero ($t_1 = 0$), it is immediate that we have $\boldsymbol{P}_1(t) = \boldsymbol{\alpha}\exp(\boldsymbol{S}t)$ where $\boldsymbol{\alpha}$ and $\boldsymbol{S}$ are the initial probability vector and the transition matrix of the phase-type distribution corresponding to the phase-type fit (Asmussen, 2003, Section III.4).

- The second patient, arriving x_1 after the first patient, enters the system when the first patient is still in the system (with probability vector $\boldsymbol{P}_1(x_1)$) or when the system is finished (with probability $F_1(x_1)$). This entails that

$$\boldsymbol{P}_2(t) = \big(\boldsymbol{P}_1(x_1), \boldsymbol{\alpha}F_1(x_1)\big)\exp(\boldsymbol{S}_2 t), \text{ for } t \geq 0$$

  with the extended transition matrix:

$$\boldsymbol{S}_2 := \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\alpha} \\ \mathbf{0}_{m,m} & \boldsymbol{S} \end{pmatrix}, \quad \text{with } \mathbf{0}_{m,m} \text{ an all-zero matrix.}$$

- The probabilities related to the i-th patient, arriving x_{i-1} after her predecessor, can be found in a similar way. To ease notation, define the matrix $\boldsymbol{T}_i$ of dimension

$(i-1)m \times m$ and $\boldsymbol{S}_i$ the transition matrix of i patients by

$$\boldsymbol{T}_i := (\mathbf{0}_{m,m}, \mathbf{0}_{m,m}, \ldots, \mathbf{0}_{m,m}, \boldsymbol{s\alpha})^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{S}_i := \begin{pmatrix} \boldsymbol{S}_{i-1} & \boldsymbol{T}_i \\ \mathbf{0}_{m,(i-1)m} & \boldsymbol{S} \end{pmatrix}.$$

Now the vector $\boldsymbol{P}_i(t)$ corresponding to patient i can be found from the vector of her predecessor $\boldsymbol{P}_{i-1}(t)$ in combination with $F_{i-1}(x_{i-1})$, by the recursion $\boldsymbol{P}_i(t) = \big(\boldsymbol{P}_{i-1}(x_{i-1}), \boldsymbol{\alpha} F_{i-1}(x_{i-1})\big) \exp(\boldsymbol{S}_i t),\ \text{for } t \geq 0$.

REFERENCES

Asmussen, S. (2003). *Applied Probability and Queues*. Stochastic Modelling and Applied Probability. New York, NY, USA: Springer-Verlag.