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 $\alpha$ whose entries sum to 1, and an $(m \times m)$-dimensional transition rate matrix $S$ whose entries satisfy $s_{ii} < 0$, $s_{ij} \geq 0$ and $\sum_{j=1}^{m} s_{ij} \leq 0$. Informally, the phase-type random variable is then the time it takes for a Markov process with rate matrix $S$ to reach an absorbing state, starting in the initial distribution $\alpha$. As can be checked easily, in the phase-type fit for $\varrho$ 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 $S$ 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 $\varrho$ 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 $p_{j,k}^{(i)}(t) = \mathbb{P}(N_i(t) = j, K_i(t) = k)$. Our objective is to evaluate the following vector of dimension $mi$:

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

In other words, the $mi$-dimensional vector $P_i(t)$ 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

\[ F_i(t) := P(S_i \leq t) = p_{0,0}^{(i)}(t) = 1 - \sum_{j=0}^{i-1} \sum_{k=1}^{m} p_{j,k}^{(i)}(t) = 1 - P_i(t) \mathbf{1}_{mi}, \]

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 \( P_i(t) \) is best explained by describing it on a patient-by-patient basis.

\( \circ \) For the first patient, arriving at time zero \( (t_1 = 0) \), it is immediate that we have \( P_1(t) = \alpha \exp(St) \) where \( \alpha \) and \( S \) 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).

\( \circ \) 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 \( P_1(x_1) \)) or when the system is finished (with probability \( F_1(x_1) \)). This entails that

\[ P_2(t) = (P_1(x_1), \alpha F_1(x_1)) \exp(S_2t), \text{ for } t \geq 0 \]

with the extended transition matrix:

\[
S_2 := \begin{pmatrix}
S & s\alpha \\
0_{m,m} & S
\end{pmatrix}, \text{ with } 0_{m,m} \text{ an all-zero matrix.}
\]

\( \circ \) 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 \( T_i \) of dimension
\((i - 1)m \times m\) and \(S_i\) the transition matrix of \(i\) patients by

\[
T_i := (0_{m,m}, 0_{m,m}, \ldots, 0_{m,m}, s\alpha)^T \quad \text{and} \quad S_i := \begin{pmatrix} S_{i-1} & T_i \\ 0 \cdot (i-1)m & S \end{pmatrix}.
\]

Now the vector \(P_i(t)\) corresponding to patient \(i\) can be found from the vector of her predecessor \(P_{i-1}(t)\) in combination with \(F_{i-1}(x_{i-1})\), by the recursion \(P_i(t) = (P_{i-1}(x_{i-1}), \alpha F_{i-1}(x_{i-1})) \exp(S_i t), \text{ for } t \geq 0.\)

**REFERENCES**

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