Phase-type Distributions in Healthcare Modelling III

Mark Fackrell

Department of Mathematics and Statistics
University of Melbourne
Victoria 3010
AUSTRALIA

email: mfackrel@ms.unimelb.eu.au

In the last of these three articles on phase-type (PH) distributions in healthcare modelling, I will suggest that more general PH distributions be considered when modelling systems. As explained in the previous article, Coxian distributions have been used predominately in healthcare modelling, mainly because of their simplicity and ability to give some sort of interpretation to the systems being modelled. However, more general PH distributions may sometimes be more useful because of their greater versatility.

Consider the histogram of some length of stay data shown in Figure 1. If an order 6 general PH distribution is fitted to the data using the EM (*Expectation-Maximization*) algorithm (see Asmussen, Nerman, and Olsson [1]) the resultant representation is

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{1}$$

$$\boldsymbol{T} = \begin{pmatrix} -3.2115 & 3.2115 & 0 & 0 & 0 & 0 \\ 0 & -3.2115 & 0 & 3.2115 & 0 & 0 \\ 0.6086 & 0 & -0.6272 & 0 & 0.0186 & 0 \\ 0 & 0 & 0 & -3.2115 & 0 & 3.2115 \\ 0 & 0 & 0.8053 & 0 & -0.8053 & 0 \\ 0 & 0 & 0 & 0 & 1.6479 & -3.2115 \end{pmatrix}. \tag{2}$$

This PH distribution cannot be a Coxian distribution (of any order) because some of the eigenvalues of $\boldsymbol{T}$ are complex numbers. The corresponding density function is also shown in Figure 1, and as we can see, the fit is quite good - the loglikelihood being -11706.9226. Using the EM algorithm, an order 25 Coxian distribution is needed to achieve a fit with a greater loglikelihood. Here, it appears, using a general PH representation is superior to using a Coxian representation. Indeed, if the representation (1)–(2) is needed in the calculation of performance measures, the smaller representation will be much easier to compute with than a larger Coxian representation. Note also that the representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ has only 5 "free" parameters, with values $3.2115, 0.6086, 0.6272, 0.8053$, and 1.6479. Recall that a general order 6 PH distribution requires 11 parameters, and with this particular example we have observed an even further reduction in the number of parameters needed. This curious observation occurs quite a lot when fitting PH distributions to data, but is not well understood. For some discussion on this aspect of PH fitting see Faddy [3] and [4], and Hampel [5].

Figure 2 shows a schematic and simplified diagram for patient flow in a hospital. Patients enter the hospital via the emergency department (ED) (state 1), or as elective patients requiring surgery in the theatre (state 2). After spending time in the ED patients can move to the theatre, or to one of the two wards (states 5 and 6). From the theatre patients go to the intensive care unit (ICU) (state 3), and then on to the high dependency

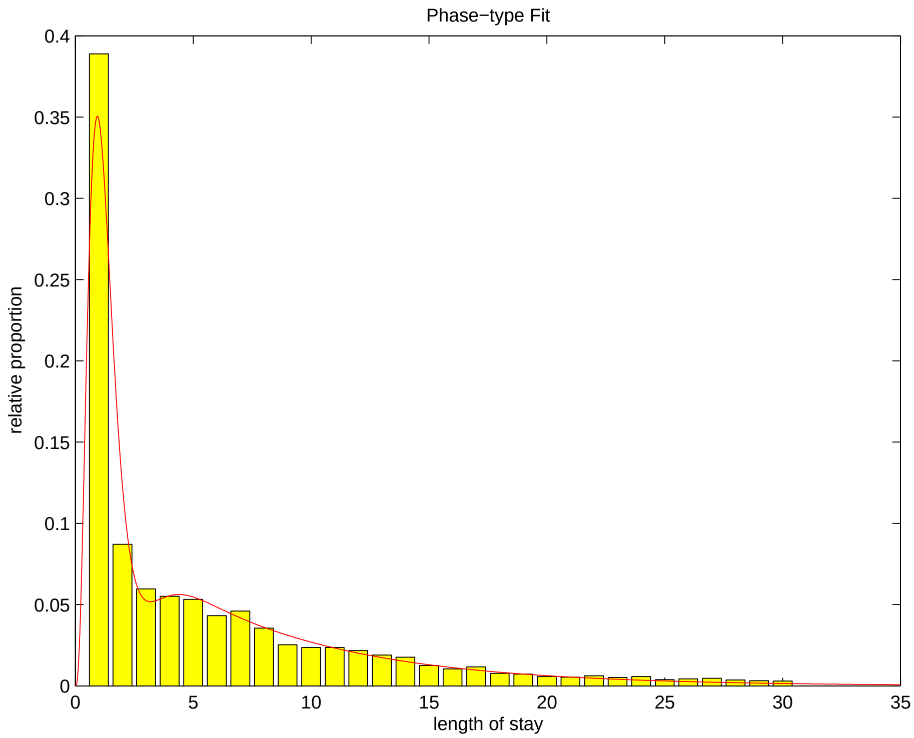

Figure 1: Order 6 *PH* fit to the length of stay histogram.

ward (HD) (state 4), before moving on to one of the two wards. At any time, patients may need to be readmitted to the ICU from HD or a ward, or they may exit the wards by being discharged or dying.

If we need to model the length of time a patient stays in hospital, we could model the length of stay in each unit with a *PH* distribution, and then combine them according to the structure shown in Figure 2 to form a larger *PH* distribution. For, $i = 1, 2, \ldots, 6$, if unit i is modelled with an order p_i *PH* distribution, such a *PH* distribution would have a representation

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{3}$$

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{15} & \boldsymbol{T}_{16} \\ \mathbf{0} & \boldsymbol{T}_{22} & \boldsymbol{T}_{23} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{33} & \boldsymbol{T}_{34} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{43} & \boldsymbol{T}_{44} & \boldsymbol{T}_{45} & \boldsymbol{T}_{46} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{53} & \mathbf{0} & \boldsymbol{T}_{55} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{63} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{66} \end{pmatrix}. \tag{4}$$

Here, $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha}_2$ will be nonnegative and nonzero vectors of lengths p_1 and p_2, respectively, and $\boldsymbol{T}_{ii}$ is an order p_i *PH* generator. The nonzero off-diagonal matrices are all nonnegative, and have size $p_i \times p_j$ whenever $(i, j) \in \{(1, 2), (1, 5), (1, 6), (2, 3), (3, 4), (4, 3), (4, 5), (4, 6), (5, 3), (6, 3)\}$.

The simplest way to fit such a *PH* distribution to data would be to assume that the time spent in each unit is exponentially distributed, and then use the *EM* algorithm to fit an order 6 PH distribution of structure (3)–(4). Alternatively, an exponential distribution could be fitted to the length of stay data for each unit individually, and when the

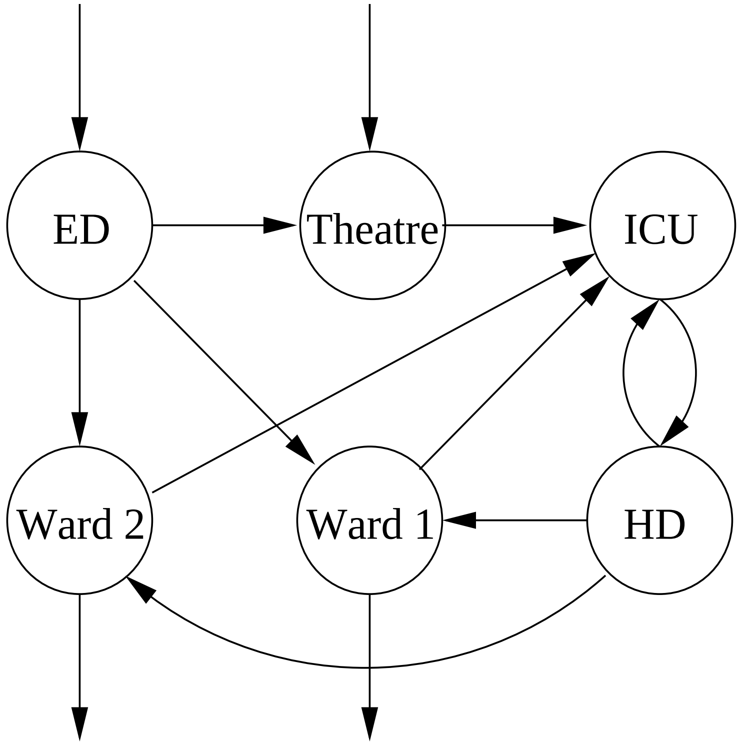

Figure 2: Schematic diagram for patient flow in a hospital.

proportions of patients moving between the units is estimated, the *PH* distribution could be constructed. This approach could be more accurate but the times patients stay in each unit needs to be recorded, rather than the total length of stay as with the former approach.

A more sophisticated approach would be to model the length of stay in each unit with a higher order *PH* (or Coxian) distribution. But here the overall representation would be quite large and the computation time taken for a good fit could be long. Also, if the time spent in each unit was fitted with a *PH* distribution individually, there is no straightforward way in which to estimate the nonzero off-diagonal matrices.

Nevertheless, to model the length of stay in this situation, general *PH* distributions would be a better choice than Coxian distributions.

A more detailed account of the use of *PH* distributions in the healthcare industry, and a comprehensive bibliography can be found in Fackrell [2].

References

[1] Asmussen, S.; Nerman, O.; Olsson, M. Fitting phase-type distributions via the *EM* algorithm. Scandinavian Journal of Statistics **1996**, *23*, 419–441.

[2] Fackrell, M. Using phase-type distributions in health and social care modelling. Submitted, 2007.

[3] Faddy, M. J. Examples of fitting structured phase-type distributions. Applied Stochastic Models and Data Analysis **1994**, *10*, 247–255.

[4] Faddy, M. J. On inferring the number of phases in a Coxian phase-type distribution. Communications in Statistics – Stochastic Models **1998**, *14*, 407–417.

[5] Hampel, K. *Modelling Phase-Type Distributions.* Honours Thesis, Department of Statistics, University of Adelaide, South Australia, 1997.