Using Prior Information on Parameters to Eliminate Dependence on Initial Values in Fitting Coxian Phase Type Distributions to Length of Stay Data in Healthcare Settings

Bin Xie[1,2,*]

[1]*Department of Obstetrics & Gyneocology;* [2]*Department of Epidemiology & Biostatistics, University of Western Ontario, Ontario, Canada*

Abstract: *Background*: Modeling length of stay (LOS) data in healthcare settings using Coxian phase type (PH) distributions is becoming increasingly popular. However, dependence on initial values is a persistent difficulty in parameter estimations. This paper explores the utility of prior information on the parameters to address this difficulty.

Methods: Maximum likelihood methods were used to estimate parameters of PH distributions that best fit simulated datasets with various sample sizes arising from PH distributions of various numbers of phases and parameters, using randomly generated initial values. Estimated values for the parameters resulting from different initial values were compared to the known values to assess the extent to which estimates depend on initial values; the impacts of sample sizes, existence of prior information, as well as the number of parameters with prior information were assessed.

Results: Without prior information, parameter estimates depend on initial values for all PH distributions and all sample sizes. Prior information on one or more parameters led to more concentrated estimates, with higher number of parameters with prior information or larger sample sizes leading to more concentrated estimates. For example, with a sample size of 500, the estimates for a parameter with known value of 0.706 without prior information had a wide range of 1.523; using prior information for two parameters narrowed that range down to 0.156. For 3-phase PH distributions, prior information on 3 parameters appeared to be sufficient to eliminate dependence on initial values, even for small sample sizes. For 4-phase PH distributions, prior information on 5 parameters and a moderate sample size were needed to eliminate such dependence.

Conclusions: Combination of prior information on parameters and sufficient sample sizes can eliminate dependence on initial values in fitting PH distributions to LOS data.

Keywords: Coxian phase type distributions, Length of stay data, Maximum likelihood methods, Prior information, Dependence on initial values.

1. INTRODUCTION

Coxian phase type distributions (PH distributions), a special type of Markov chain model that describes duration until an event occurs in terms of a process consisting of a sequence of latent phases [1-3], are becoming increasingly popular in modeling length of stay (LOS) data in healthcare settings [1-11]. Attractive properties include the ability to offer superior fit compared to alternative distributions such as lognormal or gamma distributions [4], and the ability to model probabilities of transition from one phase to another as well as probabilities of absorption from various phases [5, 6, 12].

While these properties make PH distributions an attractive modeling tool in healthcare settings, fitting PH distributions to a given dataset is not a trivial process, with some well documented difficulties including dependence on initial values and potential over-fitting or parameter redundancy [3, 13-15]. There have been numerous contributions in the literature to identify and address these issues [3, 13, 15-20], and a recent paper by Marshall and Zenga summarized what have been achieved and presented the best procedure to estimate parameters of PH distributions [3]. Despite these developments, however, no effective mechanism exists to address the heavy dependence of estimation results on initial values [3, 13].

Such dependence on initial values means that parameter values obtained from model estimations may not reliably reflect the parameters of the underlying Markov process that gave rise to the data. If this issue is not resolved, it would be difficult, if not impossible, to use the estimated parameter values to obtain information on the underlying Markov process, such as probability of transition between consecutive phases or probability of absorption in various phases. It would also be difficult, or even impossible, to reliably model dependence of parameter values of the PH distributions on covariates. Moreover, the validity and reliability of predictions using coefficient values obtained from fitting PH distribution to the data may also be questionable.

*Address corresponding to this author at the Room E2-620B, LHSC-VH, 800 Commissioners Road East, London, Ontario, N6A 5W9, Canada;
Tel: +1 519-685-8500-55174; Fax: 1-519-685-8176; E-mail: bxie5@uwo.ca

E-ISSN: 1929-6029/12

An effective mechanism to eliminate dependence on initial values, therefore, is needed to remove a significant barrier to more widespread utilization of PH distributions in healthcare settings. In this paper, I aim to establish such a mechanism. Given that methodological issues have already been explored extensively in the literature [3, 13, 15-20], I focus on the utilization of our knowledge, or prior information, on parameters of the underlying Markov process that gave rise to the data. Examples of such information or knowledge include the mean time patients spent in various phases, and the probabilities of absorption from various phases. The impact of sample sizes will also be explored, as it has been shown that larger sample sizes would result in reduced variability of the estimated parameter values and provide a more precise estimate [3].

2. METHODS

2.1. Basics of PH Distributions

A *n*-phase PH distribution represents a Markov chain with *n* transitory phases and an absorption phase, as shown in Figure **1**. Such a PH distribution has 2*n*-1 parameters, $\lambda_i, i = 1,\ldots,n-1$, and $\mu_i, i = 1,\ldots,n$. A patient's journey in this Markov chain is determined by λ_i and μ_i: the time spent on each transitory phase *i* follows an exponential distribution with mean $\frac{1}{\lambda_i + \mu_i}$, the probability of transitioning to next phase is $\frac{\lambda_i}{\lambda_i + \mu_i}$, and the probability of absorption is $\frac{\mu_i}{\lambda_i + \mu_i}$.

The transition probabilities from one state to its next state for a n-phase PH distribution as shown in Figure **1** can be written as:

$$P\{X(t+\delta t) = i+1 \,|\, X(t) = i\} = \lambda_i \delta t + o(\delta t), \text{ for } i = 1,2,\ldots,n-1$$

The probabilities of absorption can be written as:

$$P\{X(t+\delta t) = n+1 \,|\, X(t) = i\} = \mu_i \delta t + o(\delta t), \text{ for } i = 1,2,\ldots,n$$

The probability density function (pdf) of the time spent before absorption is:

$f(t) = \mathbf{p}\exp\{\mathbf{Q}t\}\mathbf{q}$, where

$\mathbf{p} = (1 \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0)$,

$$\mathbf{Q} = \begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_{n-1}+\mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -\mu_n \end{pmatrix}$$

$\mathbf{q} = (\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n)^{\mathrm{T}}$

2.2. Data Source

Figure **2** outlines the process of data generation and parameter estimation. Simulated datasets were generated from 8 PH distributions, 4 with 3 phases, and the remaining 4 with 4 phases. 3 and 4 phases were used because most published studies using PH distributions in healthcare settings used PH distributions with 3 or 4 phases. For each of these two classes of PH distributions, one set of parameter values were selected from published papers. For 3-phase PH distribution, the following parameter values were used: λ_1 = 0.0011, μ_1 = 0.0394, λ_2 = 1.56x10^{-5}, μ_2 = 0.0011, μ_3 = 0.0002 [1]. For 4-phase PH distribution, the following parameter values were used: λ_1 = 0.067, μ_1 = 0.024, λ_2 = 0.026, μ_2 = 0.065, λ_3 = 0.0011, μ_3 = 0.016, μ_4 = 0.0023 [7]. In addition to these parameter values, three sets of random numbers between 0 and 1 are also used for each class to assess the generalizabity of the study findings.

For each PH distribution, 15 datasets (3 different sample sizes, with 5 datasets per sample size per PH

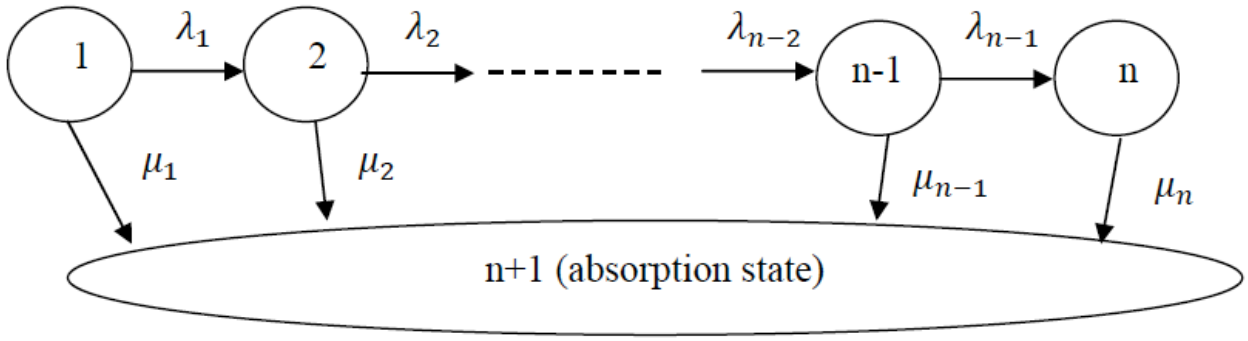

Figure 1: Schematic presentation of a Coxian phase-type distribution and its corresponding Markov chain model.

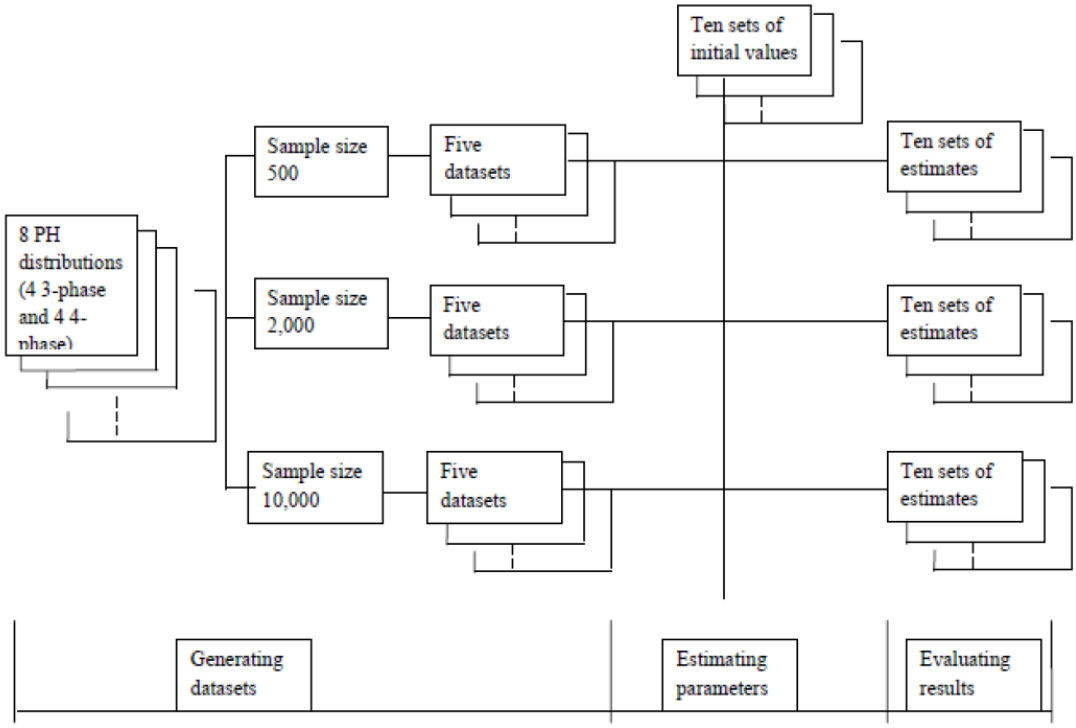

Figure 2: Illustration of data generation, parameters estimation and results evaluation.

distribution) were generated, resulting in a total of 120 datasets. The sample sizes used were 500; 2,000 and 10,000. Most published studies used sample sizes around 2,000 [1, 4, 7]; the 500 and 10,000 were included to assess the impact of different sample sizes.

2.3. Fitting PH Distributions to the Datasets

The maximum likelihood method, widely used in the literature [1, 2, 4-11] and reported as the most effective approach for fitting PH distributions [3, 13], was used in this study. The Quasi-Newton algorithm is used in the optimization, as it is shown to be more effective to compute the maximum likelihood estimates than alternatives [3]. The number of phases is determined by a sequential procedure suggested by Faddy [17]. Parameter values of the n-phase PH distributions are then estimated using maximum likelihood methods.

2.4. Selection of Initial Values

10 sets of initial values were used for each of the 120 datasets, 5 of which were within ±5% of the known parameter values, another 5 random numbers between 0 and 1.

2.5. Using Prior Information

Having prior information on a parameter was defined as knowing that the parameter was bounded by a lower and upper bounds. In the optimization, the lower and upper bounds were assumed to be 95% and 105% of the parameter, respectively. Prior information on this parameter was operationized by constraining the parameter values to be within these two boundaries in optimization. The number of parameters with prior information started with 1 (λ_1), and was increased by one a time (μ_1, then λ_2, then μ_2 etc.) until all but two parameters are constrained, given that many alternative distributions, such as lognormal and gamma distributions, have two free parameters. When fitting with prior information on certain parameters, the initial values of these parameters were random numbers within their boundaries, whereas initial values of other parameters without prior information were the same for estimations without prior information.

Starting with parameters of the first phase is due to the fact that estimates for parameters in earlier phases tend to be closer to their true values than for parameters in later phases [3], and that availability of prior information on later phases usually imply availability of prior information on the earlier phases. For example, to know information on time spent on the second phase would imply that information on time spent on the first phase is also available, as the beginning of the second phase is also the end of first phase.

2.6. Evaluation of Outcomes

For each parameter of the PH distributions, the outcome of interest is distribution of estimated values using the different combination of sample sizes, initial values, and prior information. If the all estimated values were within ±5% of the known parameter value regardless of initial values, then the combination of sample sizes and prior information was considered able to produce reliably accurate estimates of the known parameter values with no dependence on initial values.

Data generation, parameter estimation, and statistic analysis were carried out using R 2.14.0. [21].

3. RESULTS

3.1. Fitting without Prior Information

Tables **1** and **2** present the distributions of estimated parameter values for one PH distribution from each class (3-phase and 4-phase) without prior information for a particular dataset. Results of these two PH distributions with other datasets and the distributions of other PH distributions with different parameter values showed similar pattern and were not shown in the tables.

It is apparent from the tables that without prior information, the maximum likelihood methods were not able to produce estimates that are reliably close to the known values, regardless of sample sizes for most parameters. Even for sample size as large as 10,000, the estimates still depend heavily on initial values, suggesting that larger sample size is not the answer to the problem of dependence on initial values.

One exception appeared to be μ_1, as the estimates of this parameter centered around the known values regardless of initial values for both classes of PH distributions. In the case of μ_1, larger sample sizes led to more precise estimates, although the results are remarkably accurate even for smaller sample sizes. This is also consistent with results reported in Marshall and Zenga [3]. It is unclear why estimates for μ_1 did not appear to depend on initial values; it is worthwhile to explore this in further studies.

3.2. Fitting with Prior Information

Tables **3** and **4** present the distributions of estimated parameter values for one PH distribution from each class (3-phase and 4-phase) with one or more parameters constrained to lower and upper boundaries for a particular dataset. Results of these two PH distributions with other datasets andother PH distributions with different parameters showed similar pattern and were not shown in the tables.

It is apparent that constraining one or more parameters led to more concentrated estimates for other unconstrained parameters, and the more parameters that were constrained, the more concentrated the estimates of the remaining unconstrained parameters were. For PH distributions with 3 phases, constraining 3 parameters (λ_1, μ_1 and

Table 1: Range of Estimated Values of a Three-Phase PH Distribution without Prior Information Using a Single Dataset

Sample size	λ_1(TV*= 0.414)	μ_1 (TV=0.442)	λ_2 (TV=0.159)	μ_2 (TV=0.706)	μ_3 (TV=0.498)
500	0.140-0.826	0.385-0.412	0.00001-0.426	0.767-2.290	0.045-0.460
2,000	0.159-1.020	0.416-0.423	0.077-1.073	0.553-1.026	0.590-1.492
10,000	0.124-0.418	0.444-0.448	0.087-0.926	0.709-1.162	0.400-2.062

*TV means true value.

Table 2: Range of Estimated Values of a Four-Phase PH Distribution without Prior Information Using a Single Dataset

Sample size	λ_1 (TV*= 0.469)	μ_1 (TV=0.909)	λ_2 (TV=0.518)	μ_2 (TV=0.955)	λ_3 (TV=0.460)	μ_3 (TV=00.817)	μ_4 (TV=0.894)
500	0.00001-3.89	0.885-1.161	0.00001-3.010	0.0006-1.691	0.00001-2.803	0.0001-1.488	0.045-2.761
2,000	0.00001-2.41	0.809-0.940	0.030-3.021	0.135-1.402	0.00002-2.052	0.132-2.328	0.093-0.947
10,000	0.009-0.535	0.895-0.917	0.00003-0.731	0.182-1.320	0.0008-0.658	0.520-0.909	0.222-0.877

*TV means true value.

λ_1) was sufficient to eliminate dependence on initial values for all parameters, even for sample size as small as 500, as shown in Table **3**. This is true for larger sample sizes, with larger sample size leading to even more concentrated estimates (not shown in the table). For PH distributions with 4 phases, it is necessary to constrain 5 parameters (λ_1, μ_1, λ_2, μ_2, and λ_3), and to have sufficiently large sample size (2,000), to eliminate such dependence. With smaller size (i.e., 500), even constraining 5 parameters would be insufficient to eliminate dependence on initial values for the remaining two parameters, and with larger sample size (i.e., 10,000) constraining 5 produced more concentrated estimates, although constraining 4 parameters was insufficient to eliminate dependence on initial values (not shown in the tables).

Another striking result is that for both classes of PH distributions, even when estimates did not depend on initial values as a result of using prior information, the known parameter values may not fall within ±5% of the estimated values. In other words, the estimated parameter values may cluster closely around some values regardless of initial values used, suggesting that dependence on initial values were eliminated. However, these "some values" around which the estimates clustered may be different than the known parameter values.

Upon closer look, it became clear that when the estimated values cluster around values other than the know parameter values, the likelihood of the dataset given the PH distributions with the estimated parameters were higher than that of the dataset given the PH distribution with known parameter values, as shown in Table **5**. In other words, the PH distributions with the estimated parameter values fitted the dataset better than the theoretical PH distribution (the PH distribution with known parameter values), suggesting that the maximum likelihood methods worked as it is supposed to. This can happen even for large sample sizes, suggesting that larger sample size is unlikely to be the answer to avoid this problem.

4. EXAMPLE

Here I present an example on how the findings of this study can be used in practice. Data for this example came from administrative emergency department (ED) records from an adult emergency department (ED) in Canada between April 1, 2009 and July 31, 2009. The ED records include the timing of patient triage, initial consultation with nurse, initial consultation with doctor, and discharge. Details of the data source can be found elsewhere [22].

The ED process is a well defined process in which patients flow through a series of transitory phases,

Table 3: Range of Estimated Values of a Three-Phase PH Distribution with Prior Information Using a Single Dataset, Sample Size 500

Parameters constrained	λ_1(TV*= 0.414)	μ_1 (TV=0.442)	λ_2 (TV=0.159)	μ_2 (TV=0.706)	μ_3 (TV=0.498)
λ_1	0.434-0.435	0.391-0.3923	0.489-0.5019	1.073-1.081	0.434-0.437
λ_1, μ_1	0.401-0.434	0.420-0.421	0.435-0.461	0.926-0.978	0.435-0.441
$\lambda_1, \mu_1, \lambda_2$	0.405-0.434	0.420-0.421	0.156-0.166	0.659-0.815	0.480-0.522

*TV means true value.

Table 4: Range of Estimated Values of a Four-Phase PH Distribution with Prior Information Using a Single Dataset, Sample Size 2,000

Parameters constrained	λ_1 (TV*= 0.469)	μ_1 (TV=0.909)	λ_2 (TV=0.518)	μ_2 (TV=0.955)	λ_3 (TV=0.460)	μ_3 (TV=0.817)	μ_4 (TV=0.894)
λ_1	0.449- 0.492	0.864-0.868	0.00007-3.341	0.0004-1.044	0.00001-2.803	0.120-2.158	0.185-0.909
λ_1, μ_1	0.446-0.492	0.863-0.864	0.493-0.544	0.663-0.999	0.0007-0.543	1.076-3.759	0.129-0.721
$\lambda_1, \mu_1, \lambda_2$	0.446-0.491	0.863-0.864	0.496-0.541	0.907-0.988	0.0002-0.909	1.064-1.807	0.731-1.880
$\lambda_1, \mu_1, \lambda_2, \mu_2$	0.445-0.492	0.863-0.864	0.494-0.542	0.907-1.000	0.438-0.471	0.880-0.916	0.772-0.789
$\lambda_1, \mu_1, \lambda_2, \mu_2, \lambda_3$	0.484-0.492	0.863-0.864	0.492-0.497	1.002-1.003	0.437-0.480	0.856-0.857	1.421-1.455

*TV means true value.

Table 5: Loglikelihood Values

Distribution	Using known values	Using estimated values
3-phase, with prior information on 3 parameters, sample size 500 (as shown in Table **3**)	-825.80	-825.12 to -825.11
4-phase, with prior information on 5 parameters, sample size 2000 (as shown in Table **4**)	-2205.8	-2,201.86 to -2,202.17

ending with an absorption phase (i.e., discharge) [23]. As such, PH distributions could be used to fit length of stay (LOS) data at the ED. Here I present results of fitting the ED LOS data with PH distributions using findings from this paper.

Using the procedure described in section 2.3 above, it is determined that a 4-phase PH distribution is sufficient to fit the LOS data. While the data do not provide information regarding the meaning of these phases, it is reasonable to assume, based on a widely used conceptual framework of the ED process [23], that these four phases correspond to waiting to be assessed, initial examination by ED staff, further diagnostic testing, and waiting for discharge (e.g., boarding, or waiting for testing results).

Prior information that was used to estimate the parameters was obtained from the length of stay in each of the transitory phases and the transition probabilities across these phases and probabilities of absorption from some of these transitory phases. These probabilities were estimated using the proportions of patients transiting from one phase to the next and the proportions of patients discharged from each of these phases. For example, mean LOS at phase 1 was estimated by averaging the time patients spent on the waiting room (time between triage and initial consult with ED staff); transition probability from initial consult with ED staff to further testing was estimated by the proportion of patients who received further testing. Using such information, the initial values were: λ_1 =0.023, μ_1=0, λ_2=0.035, μ_2=0.023, λ_3=0.016, μ_3=0.17.

Table **6** below presents the results of the example, including parameter estimates and log-likelihood value using initial values based on prior information, and the range of estimated values of the parameters the log-likelihood values using random initial values (without prior information). It is clear that the estimates of the parameters of the PH model were close to the prior information, and that the log-likelihood value from the estimates using prior information was close to the maximum value achieved using a random initial values, providing evidence to suggest that the estimated model is a good representation of the underlying ED process.

5. DISCUSSION

The results suggest that dependence on initial values can be effectively avoided by the combination of using prior information on some parameters (3 for 3-phase PH distributions and 5 for 4-phase distributions) and sufficiently large sample sizes (500 was sufficient for 3-phase distributions, whereas 2,000 was needed for 4-phase distributions). While direct information on λ_i and μ_i may not be readily available in many healthcare settings, information on mean time spent on and the portion of patients who were absorbed from phase *i* can be collected in many situations and can be used to infer information on λ_i and μ_i, as mean time spent on and the probability of absorption from phase *i* are $\frac{1}{\lambda_i + \mu_i}$ and $\frac{\mu_i}{\lambda_i + \mu_i}$, respectively.

Prior information for a parameter was modeled as constraining the parameter to a lower and upper bound in the optimization process. In particular, 95% and 105% of the known value were used as the two boundaries. In practice, the known values are unknown, so 95% and 105% of the parameter values calculated from mean time spent on and the portion of patients who were absorbed from phase *i* can be used as the boundaries.

Table 6: Results of the Example (Sample Size:15411)

	λ_1	μ_1	λ_2	μ_2	λ_3	μ_3	μ_4	Log-likelihood
w/ prioir info	0.023	0.0004	0.036	0.023	0.015	0.17	0.020	-156889.6
w/o prior info*	0.0001-1.65	0.0001-0.68	0.005-0.54	0.0002-1.02	0.0001-0.68	0.0003-0.48	0.01-0.62	-156860.1--158619.8

*1000 runs using random numbers between 0 and 1 as initial values.

The results also suggest that even after the combination of prior information and sufficient sample size eliminated dependence on initial values, the known parameter values may not fall within ±5% of the estimated values. This is due to the fact that some PH distributions with different sets of parameter values have higher likelihood of giving rise to the dataset than the PH distribution corresponding to the underlying Markov process.

The fact that the some distributions with parameter values that differ with the known values fit the given dataset better than the theoretical distribution that gave rise to the dataset is not unique to PH distributions. It is a consequence of the randomness of the data and can happen in fitting any distributions to a given dataset. For simpler distributions with fewer parameters (e.g., normal or gamma distributions), however, the estimated values will be close to the known values, and with sufficient sample size, the known values will be within ±5% of estimated values. This is however not necessarily so for PH distributions due to the non-linear nature of the likelihood function for PH distributions. This cannot be resolved by using a larger sample size. Even more parameters would need to be constrained to avoid this problem.

Results of this study also suggest that larger sample sizes do not always help in fitting PH distributions to a given dataset. While larger sample sizes appear to provide more concentrated estimates, as suggested in the literature [3], without prior information, even sample size as large as 10,000 was unable to avoid dependence on initial values. With prior information, a sample size of 2,000 was sufficient to produce impressively clustered estimates regardless of initial values; the utility of larger sample sizes is very limited. Furthermore, large sample size is not helpful in addressing the issue of residual non-linearity for PH distributions. Given the cost of collecting larger sample sizes, results of this study suggest that it may make more sense to collect prior information on the underlying Markov process than to collect larger sample sizes.

Based on these findings, it appears reasonable to suggest that when using PH distributions to model LOS data in healthcare settings, it is important to have a clear plan to address the dependence on initial values. When such plan involves collecting prior information on parameters, the costs of obtaining such prior information to ensure reliable estimates should be weighed against the potential gain of using PH distributions instead of simpler distributions such as gamma or lognormal distributions. Only when the potential gain outweigh the costs should modeling with PH distributions be used.

There are a number of limitations that need to be acknowledged. For one thing, this paper did not provide clear theoretical explanations to some of the findings, in particular the fact that estimates of μ_1. were always clustered around the known value. Another limitation is that this paper did not provide a mechanism to estimate the sample size required to produce a pre-defined level of accuracy. Another limitation is that this study did not address the issue of modeling dependence of parameter values on covariates. In what situations can such dependence be reliably modeled and coefficients be reliably and accurately estimated? Expanding the research to addressing these limitations are fruitful areas of further research.

6. Conclusion

The combination of prior information on parameters and sufficient sample size can successfully eliminate of the dependence on initial values in fitting PH distributions to LOS data. In practice, the costs of obtaining such prior information and sample size should be weighed against the gain of using PH distributions compared with simpler distributions.

REFERENCES

[1] Marshall AH, McClean SI. Using Coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. Health Care Manag Sci 2004; 7(4): 285-89. http://dx.doi.org/10.1007/s10729-004-7537-z

[2] Aaronson LS, Mural CM, Pfoutz SK. Seeking information: where do pregnant women go? Health Educ Q 1988 Fall; 15(3): 335-45. http://dx.doi.org/10.1177/109019818801500307

[3] Marshall AH, Zenga M. Experimenting with the Coxian Phase-Type Distribution to Uncover Suitable Fits. Methodology and computing in applied probability 2012; 14(1): 71-86. http://dx.doi.org/10.1007/s11009-010-9174-y

[4] Faddy M, Graves N, Pettitt A. Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions. Value Health 2009; 12(2): 309-14. http://dx.doi.org/10.1111/j.1524-4733.2008.00421.x

[5] Faddy MJ, McClean SI. Using a multi-state model to enhance understanding of geriatric patient care. Aust Health Rev 2007; 31(1): 91-97. http://dx.doi.org/10.1071/AH070091

[6] Faddy MJ, McClean SI. Markov chain modelling for geriatric patient care. Methods Inf Med 2005; 44(3): 369-73.

[7] Faddy MJ, McClean SI. Analysing data on lengths of stay of hospital patients using phase-type distributions. Appl Stochastic Models Bus Ind 1999; 15(4): 311-17. http://dx.doi.org/10.1002/(SICI)1526-4025(199910/12)15:4<311::AID-ASMB395>3.0.CO;2-S

[8] Marshall AH, Shaw B, McClean SI. Estimating the costs for a group of geriatric patients using the Coxian phase-type distribution. Stat Med 2007; 26(13): 2716-29. http://dx.doi.org/10.1002/sim.2728

[9] Marshall AH, McClean SI, Shapcott CM, Millard PH. Modelling patient duration of stay to facilitate resource management of geriatric hospitals. Health Care Manag Sci 2002; 5(4): 313-19. http://dx.doi.org/10.1023/A:1020394525938

[10] Marshall AH, McClean SI, Millard PH. Addressing bed costs for the elderly: a new methodology for modelling patient outcomes and length of stay. Health Care Manag Sci 2004; 7(1): 27-33. http://dx.doi.org/10.1023/B:HCMS.0000005395.77308.d1

[11] Xie H, Chaussalet TJ, Millard PH. A continuous time Markov model for the length of stay of elderly people in institutional long-term care. J Royal Statist Soc 2005; 168: 51-61. http://dx.doi.org/10.1111/j.1467-985X.2004.00335.x

[12] Marshall AH, McClean SI. Using Coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. Health Care Manag Sci 2004; 7(4): 285-89. http://dx.doi.org/10.1007/s10729-004-7537-z

[13] Marshall AH, Zenga M. Simulating Coxian phase-type distributions for patient survival. Int Trans Operat Res 2009; 16(2): 213-26. http://dx.doi.org/10.1111/j.1475-3995.2009.00672.x

[14] Asmussen S, Nerman O, Olsson M. Fitting Phase-Type Distributions *via* the EM Algorithm. Scand J Stat 1996; 23(4): 419-41.

[15] Lang A, Arthur JL. Parameter approximation for phase-type distributions. In: Chakravarthy SR, Alfa AS, Eds. Matrix-analytic methods in stochastic models (Lecture notes inpure and applied mathematics). New York: Marcel Dekkar 1996; pp. 151-206.

[16] Augustin R, Buscher KJ. Characteristics of the Cox-Distribution. Perform Eval Rev 1982; 12(1): 22-32. http://dx.doi.org/10.1145/1041818.1041821

[17] Faddy M. On inferring the Number of Phases in a Coxian Phase-Type Distribution. Commun Statistics-Stochastic Models 1998; 14: 407-17. http://dx.doi.org/10.1080/15326349808807479

[18] Faddy M. Examples of fitting structured phase-type distributions. Appl Stoch Model Data Anal 1994; 10: 247-55. http://dx.doi.org/10.1002/asm.3150100403

[19] Faddy MJ. Penalized maximum likelihood estimation of the parameters in a Coxian phase-typedistribution. In: Latouche G, Taylor P, editors. Matrix-analytic Methods. Theory and Applications. New Jersey: World Scientific 2002; pp. 107-114.

[20] Johnson MA. Selecting parameters of phase distribution: combining non linear programming,heuristics and Erlang distribution. ORSA J Comput 1993; 5: 69-83. http://dx.doi.org/10.1287/ijoc.5.1.69

[21] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2012.

[22] Xie B, Youash S. The effects of publishing emergency department wait time on patient utilization patterns in a community with two emergency department sites: a retrospective, quasi-experiment design. Int J Emerg Med 2011; 4(1): 29. http://dx.doi.org/10.1186/1865-1380-4-29

[23] Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA Jr. A conceptual model of emergency department crowding. Ann Emerg Med 2003; 42(2): 173-80. http://dx.doi.org/10.1067/mem.2003.302

Received on 28-09-2012 Accepted on 29-11-2012 Published on 21-12-2012

http://dx.doi.org/10.6000/1929-6029.2012.01.02.02