Parametric landmark estimation of the transition probabilities in survival data with multiple events

GUSTAVO SOUTINHO
EPIUnit
University of Porto
Rua das Taipas 135, 4050-600 Porto
PORTUGAL

LUÍS MEIRA-MACHADO
University of Minho
Centre of Mathematics
Campus de Azurém, 4800-058 Guimarães
PORTUGAL

Abstract: Multi-state models are a useful tool for analyzing survival data with multiple events. The transition probabilities play an important role in these models since they allow for long-term predictions of the process in a simple and summarized manner. Recent papers have used the idea of subsampling to estimate these quantities, providing estimators with superior performance in the case of strong violations of the Markov condition. Subsampling, also referred to as landmarking, leads to small sample sizes and usually heavily censored data, which leads to estimators with higher variability. Here, we use the flexibility of the generalized gamma distribution combined with the same idea of subsampling to obtain estimators free of the Markov condition with less variability. Simulation studies show the good small sample properties of the proposed estimators. The proposed methods are illustrated using real data.

Key-Words: Multi-state models, Transition probabilities, Generalized gamma distribution, Landmark approach

Received: June 6, 2019. Revised: March 12, 2022. Accepted: April 2, 2022. Published: May 3, 2022.

1 Introduction

Multi-state models are models for a stochastic process that at any time occupies one of a set of discrete states [1, 2, 3, 4]. These models provide a relevant modeling framework to deal with complex survival data in which individuals may experience more than one event. A multi-state model can be represented schematically by diagrams with boxes representing the states and arrows representing the possible transitions. The complexity of a multi-state model greatly depends on the number of states and the transitions allowed between these states.

Among the examples, the simplest case is the mortality model for survival data, which involves only two states and one transition. The competing risks model [5, 6] can be seen as an extension of the mortality model, considering that a subject may reach the ultimate state due to any of several causes. The irreversible illness-death or disability model is a special case of multi-state model, commonly used in the literature to introduce the theoretical background of multi-state models, in which the individuals may pass from the initial state (State 1) to the intermediate state (State 2) and then to the absorbing state (State 3) (Figure 1). Individuals are at risk of death in each transient state (States 1 and 2).

One important goal in the analysis of multi-state survival data is to assess the relationship of explanatory variables to the several event times. Another primary goal is to obtain predictions of the clinical prognosis of a patient at a certain point in his or her illness process. The main assumption implied in both cases is the Markov assumption, which claims that given the present state, the future evolution of the process is independent of the states previously visited and the transition times among them. Traditionally, the transition probabilities are estimated using the Aalen-Johansen estimator [7] that gives consistent estimators of the transition probabilities when the multi-state model is Markov. When the multi-state model is non-Markov, this is no longer the case. Recently, [8] and [9] introduced alternative estimators based on subsampling (also known as landmarking) that are consistent regardless of the Markov assumption. Both methods are derived from a subset of the data consisting of all subjects observed to be in the given state at the given time. The first paper, by [8], uses a procedure based on (differences between) Kaplan–Meier estimators, whereas the second, by [9] used the Aalen-Johansen estimate of the state occupation probabilities derived from the same subset. Since the computation of the transition probabilities of the two methods is based on reduced data, they will lead to estimators with higher variability. To overcome this issue, parametric estimation based on the same specific portions of data may be used to provide estimators with less variability that are free of the Markov condition. In some situations, fully-parametric modeling can be more convenient to model complex data structures and processes [10, 11]. In this paper, we use

the flexibility of the generalized gamma distribution combined with the same idea of subsampling to obtain estimators that are free of the Markov condition with less variability.

The following section provides some notation and introduces the theoretical background associated with the landmark estimators and the proposed estimators based on the generalized gamma distribution. In section 3, these methods are compared through simulations. The application to a real data set on colon cancer is presented in Section 4. Finally, the main conclusions of this paper are reported in Section 5.

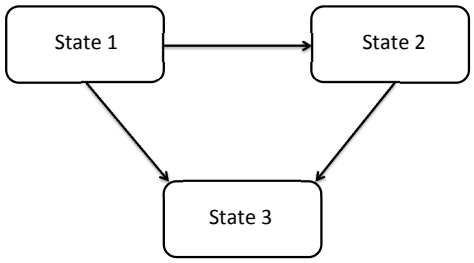

Figure 1: The progressive illness-death model.

2 Models and methods

2.1 Notation

In the following, we will devote our attention to the particular case of the illness–death model that involves three different states and three possible transitions among them, since, as mentioned previously, this model plays an important role in the theory of multi-state models. The process of the illness-death model is fully characterized by the vector of times (Z, T), where Z denotes the sojourn time in the initial state (State 1) and T the total survival time (i.e., the absorption time). Note that for those individuals undergoing a direct transition from State 1 to the absorbing State 3 we have $Z = T$. On the other hand, $Z < T$ indicates that the individual visits the intermediate State 2 at some time. We assume that right-censoring issues may appear due to time limitations in the follow-up, lost follow-up cases, etc. This censoring is modeled by considering a censoring variable, C, which we assume to be independent of the process. Under censoring, only the censored versions of Z and T, along with their corresponding censoring indicators, are available. Thus, we define $\widetilde{Z} = \min(Z, C)$ and $\widetilde{T} = \min(T, C)$ for the censored versions of Z and T and introduce $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$ for the respective censoring indicators of Z and T. The variables $\widetilde{Z}$ and $\widetilde{T}_{23} = \widetilde{T} - \widetilde{Z}$ are the observed sojourn times in states 1 and 2, respectively. Finally, the available data are $(\widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i)$, $1 \leq i \leq n$, i.i.d. copies of $(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta)$.

2.2 Nonparametric landmark estimators of the transition probabilities

The target is each of the five transition probabilities $p_{hj}(s,t) = P(X(t) = j | X(s) = h)$, where $1 \leq h \leq j \leq 3$ and s and t are two pre-specified time points with $s < t$. As noted by [8], these transition probabilities are functions involving Z and T as follow:

$$
\begin{aligned}
p_{11}(s,t) &= P(Z > t \mid Z > s), \\
p_{12}(s,t) &= P(Z \leq t, T > t \mid Z > s), \\
p_{13}(s,t) &= P(T \leq t \mid Z > s), \\
p_{22}(s,t) &= P(Z \leq t, T > t \mid Z \leq s, T > s), \\
p_{23}(s,t) &= P(T \leq t \mid Z \leq s, T > s).
\end{aligned}
$$

Two obvious relations stand out: $p_{11}(s,t) + p_{12}(s,t) + p_{13}(s,t) = 1$ and $p_{22}(s,t) + p_{23}(s,t) = 1$, revealing that in practice one only need to estimate three of these quantities.

When the multi-state model is Markovian, the Aalen-Johansen estimator [7] gives consistent estimators of the transition probabilities. Explicit formulae of the Aalen-Johansen estimator for the illness-death model are available [13]. When the multi-state model is non-Markov, this is no longer the case. In these cases, the Aalen-Johansen estimator may introduce some bias, and therefore, they may be inappropriate.

Estimators that do not rely on the Markov assumption were first introduced by [14]. The proposed estimators were defined in terms of multivariate Kaplan-Meier integrals and proved to be more efficient than the Aalen-Johansen estimators in the case of strong violations of the Markov assumption. However, this proposal has the drawback of requiring that the support of the censoring distribution contain the support of the lifetime distribution, otherwise they only report valid estimators for truncated transition probabilities. Recently, [8] recovered the work by [14] and proposed alternative estimators that are consistent regardless of the Markov condition and the referred assumption on the censoring support. The idea of the new methods is to use a procedure based on differences between Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in a given state at a given time. Following equations above, given the time point s, to estimate $p_{1j}(s,t)$ for $j = 1, 2, 3$ the landmark analysis is restricted to the individuals observed in State 1 at time s. For the subpopulation $Z > s$, the censoring time C is still independent of the pair (Z, T)

and, therefore, Kaplan-Meier-based estimation will be consistent. Similarly, to estimate $p_{2j}(s,t)$, $j = 2,3$, the landmark analysis proceeds from the sample restricted to the individuals observed in State 2 at time s. Then, we may formally introduce the landmark estimators as follows:

$$\widehat{p}_{11}^{\,\mathtt{LM}}(s,t) = \widehat{S}_1^{\,\mathtt{KM}(s)}(t),$$

$$\widehat{p}_{12}^{\,\mathtt{LM}}(s,t) = \widehat{S}^{\,\mathtt{KM}(s)}(t) - \widehat{S}_1^{\,\mathtt{KM}(s)}(t),$$

$$\widehat{p}_{13}^{\,\mathtt{LM}}(s,t) = 1 - \widehat{S}^{\,\mathtt{KM}(s)}(t),$$

$$\widehat{p}_{22}^{\,\mathtt{LM}}(s,t) = \widehat{S}^{\,\mathtt{KM}[s]}(t),$$

$$\widehat{p}_{23}^{\,\mathtt{LM}}(s,t) = 1 - \widehat{S}^{\,\mathtt{KM}[s]}(t)$$

where $\widehat{S}_1^{\,\mathtt{KM}(s)}$ and $\widehat{S}^{\,\mathtt{KM}(s)}$ are the Kaplan-Meier estimators for the distributions of Z and T, respectively, but computed from the subsample $\mathcal{S}_1 = \left\{i : \widetilde{Z}_i > s\right\}$; whereas $\widehat{S}^{\,\mathtt{KM}[s]}$ is the Kaplan-Meier estimator of the distribution of T but computed from the subsample $\mathcal{S}_2 = \left\{i : \widetilde{Z}_i \leq s < \widetilde{T}_i\right\}$.

2.3 Parametric estimation

Nonparametric estimators of the transition probabilities based on landmarking were introduced in the previous section. However, methods based on a fully-specified model can be more convenient for representing complex data structures and processes [10]. This is the case for hazard functions that may vary predictably due to interval censoring, frailties, or multiple responses. In fact, the use of parametric models is an attractive option because standard methods, such as maximum likelihood, are available for parameter estimation and testing, and a complete description of the hazard function can be obtained [15]. [16] also refers an example in biomedical application for which the mean survival needs the survival function $S(t)$ to be fully-specified for all times t, and parametric models that combine data from multiple time periods should be used. Nevertheless, since the standard parametric distribution has hazard functions that can totally change shape according to the values of the parameters, the choice of the most suitable model from the available families of distributions can be problematic, and the results of the inference may be affected [17]. The use of parametric models may also improve the accuracy of the transition probabilities by reducing the high variability of the landmark estimators for cases with small sample sizes or high censoring levels.

The generalized gamma (GG) distribution is a very flexible distribution that generalizes several important and common parametric models that have wide applications in lifetime data analysis. Among these are the exponential, Weibull, log-normal, and two-parameter gamma distributions. In fact, due to the flexible variety of shapes of its hazard function, which includes all four of the most common types of hazard function (monotonically increasing and decreasing, as well as bathtub and arc-shaped hazards), the GG distribution can easily be used in different contexts for modelling lifetime models [18]. Due to its importance, the following survival and density functions for the GG distribution are introduced. Further details on other special distributions can be found in [18].

2.3.1 The generalized gamma distribution

The generalized gamma distribution (GG), introduced by [19], is a three-parameter distribution with the following probability density function

$$f\left(t \mid \beta > 0, \alpha > 0, k > 0\right) = \\ = \frac{\beta}{\Gamma\left(k\right)} \cdot \frac{t^{\beta k - 1}}{\alpha^{\beta k}} \cdot \exp\left[-\left(t/\alpha\right)^{\beta}\right], t > 0, \qquad (1)$$

This distribution is a generalization of the two-parameter gamma distribution when $\beta = 1$. The widely used exponential ($\beta = k = 1$) and weibull ($k = 1$), are also special cases of the GG distribution. [20] considers a new reparametrization of the GG distribution with location ($\mu = \log(\alpha)$), scale ($\sigma = \beta^{-1}$) and shape ($\lambda = 1/k^2$) that allows to extend the original one to include a further class of models with $\lambda < 0$. Specifically, let $\Gamma\left(t;\gamma\right) = \int_0^t x^{\gamma-1}e^{-x}dx/\Gamma(\gamma)$ the gamma distribution with mean and variance equal to $\gamma > 0$. The survival function for $\mathrm{GG}(\mu,\sigma,\lambda)$, if $\lambda > 0$, is given by

$$S\left(t\right) = 1 - \Gamma\left[\lambda^{-2}\exp\left(\lambda\left[log(t) - \mu\right]/\sigma\right);\lambda^{-2}\right] \qquad (2)$$

$$= 1 - \Gamma\left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma};\lambda^{-2}\right] \qquad (3)$$

and, if $\lambda < 0$, replacing $z = (log(t) - \mu)/\sigma$ by $-z$, take the following form

$$S\left(t\right) = \Gamma\left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma};\lambda^{-2}\right] \qquad (4)$$

Finally, the probability density function is

$$f\left(t\right) = \frac{|\lambda|}{\sigma t\Gamma\left(\lambda^{-2}\right)} \times \\ \times \left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right]^{\lambda^{-2}} \exp\left[-\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right] \qquad (5)$$

and the hazard function, for $\lambda > 0$, is given by

$$h(t) = \frac{f(t)}{S(t)} = \frac{|\lambda|}{\sigma t \Gamma(\lambda^{-2})} \left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right]^{\lambda^{-2}} \times \exp\left[-\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right] \times \left\{1 - \Gamma\left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}; \lambda^{-2}\right]\right\}^{-1} \quad (6)$$

and for $\lambda < 0$, has the form

$$h(t) = \frac{|\lambda|}{\sigma t \Gamma(\lambda^{-2})} \left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right]^{\lambda^{-2}} \times \exp\left[-\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}\right] \times \left\{\Gamma\left[\lambda^{-2}\left(e^{-\mu}t\right)^{\lambda/\sigma}; \lambda^{-2}\right]\right\}^{-1} \quad (7)$$

Figures 2, 3 and 4 plot the GG density, survival and hazard functions for different choices of the parameters μ, σ and λ.

In case of $\lambda = \sigma$ the GG distribution becomes the two-parameter gamma distribution $\mathrm{G}(\mu, \sigma)$. The density for the weibull distribution can be obtained with $\lambda = 1$ and the exponential distribution corresponds to the case $\lambda = \sigma = 1$. Another special case is when $\lambda = 0$ which correspond to the log-normal distribution, with $S(t) = 1 - \Phi\left(\log\left(e^{-\mu}t\right)^{1/\sigma}\right)$, where $\Phi(\cdot)$ is the standard normal cumulative density function. Therefore, the GG family contains almost all the most commonly used parametric distributions [16, 17].

2.3.2 Maximum likelihood estimation

Let's consider the probability density function of a general parameter model, for an event at time t, given by $f(t|\mu, \alpha)$, $t \geq 0$, where μ correspond to the mean or *location* parameter and the vector $\alpha = (\alpha_1, \cdots, \alpha_R)$ represents other parameters such as shape, variance or higher moments. A standard approach for the estimation of parameters of a fully parametric model is to use the likelihood function. Thus, let t_i, $i = 1, \ldots, n$ be a sample of times from individuals i, with $c_i = 1$ if t_i is an observed time-to-event, or $c_i = 0$ if this is censored. Also let s_i be corresponding left-truncation (or delayed-entry) times, meaning that the ith survival time is only observed conditionally on the individual's having survived up to s_i, thus $s_i = 0$ if there is no left-truncation. In the case of right-censoring, the likelihood of the probability of density function for the previous parameters $\theta = (\mu, \alpha)$ is [16]:

$$l(\theta|t, c, s) = \frac{\prod_{i:c_i=1} f_i(t_i|\mu, \alpha) \prod_{i:c_i=0} S_i(t_i|\mu, \alpha)}{\prod_i S_i(s_i)} \quad (8)$$

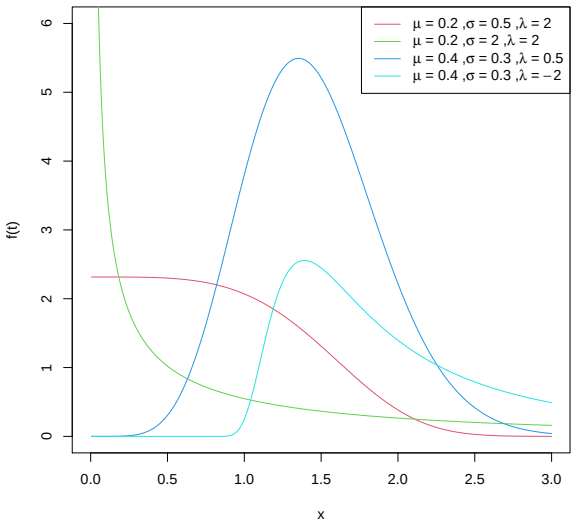

Figure 2: Plots of the GG density for some values of μ, σ and λ parameters.

In the case of the GG distribution, the likelihood function is obtained by replacing, respectively, the density and survival functions by the expressions (5) and (3 or 4, according to the value of λ). The log-likelihood function is given from (6) as follow [16]:

$$\log l(\theta|t, c, s) = \sum_{i:c_i=1} \{\log(h_i(t_i) - H_i(t_i))\} - \sum_{i:c_i=0} H_i(t_i) + \sum_i H_i(s_i) \quad (9)$$

where H denote the cumulative hazard function.

The parameters can be then estimated by maximizing the log-likelihood function with respect to θ using the maximum likelihood estimation (MLE) method. If the likelihood function is differentiable, the derivative test for determining the maximum can be applied. In some cases, the first-order conditions of the likelihood function can be solved explicitly. However, in most cases, numerical methods will be necessary to find the maximum of the likelihood function [18]. In this paper, the maximization was done using the `flexsurvreg` package, a software application for R, through the optimization methods available in the *optim* function.

2.3.3 Parametric estimators of the transition probabilities

The parametric landmark estimators are based on the landmark ideas as used in (1). This estimator can be

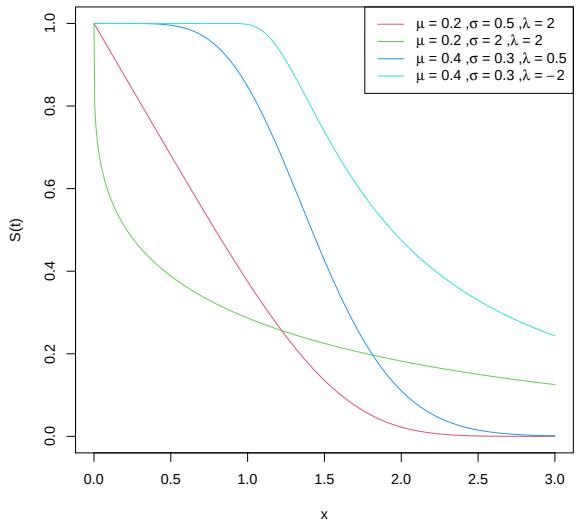

Figure 3: Plots of the GG survival functions for some values of μ, σ and λ parameters.

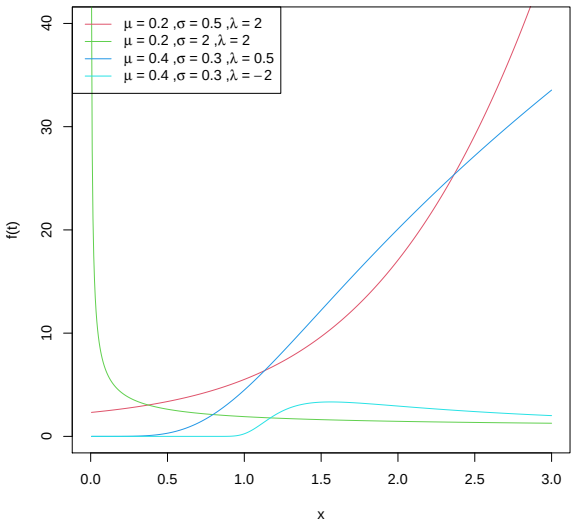

Figure 4: Plots of the GG hazard functions for some values of μ, σ and λ parameters.

obtained by applying the procedure based on (differences between) parametric survival estimators introduced in Section 2.3.1 to the same specific portions of data.

3 Simulation studies

In this section, we report the results from simulation in order to evaluate the performance of the proposed parametric estimators. In particular, we compare the estimator described in Section 2.3, using the GG distribution (labeled GG, to the Landmark estimator introduced by [8], labeled LM). Data were simulated to represent a progressive illness-death model, assuming that all individuals begin in the initial state (State 1) at time $t = 0$. The movement of the individuals was defined according to the scenario described by [21] and [22] in which the subjects may pass through the intermediate state (State 2) at some time or go directly to the absorbing state (State 3). For the individuals who visit the intermediate state, we generated replicates of $(Z, T - Z)$ through the bivariate distribution $F_{12}(x, y) = F_1(x)F_2(y)\left[1 + \{1 - F_1(x)\}\{1 - F_2(y)\}\right]$ with exponential marginal distribution functions with rate parameter 1. For individuals that experience a direct transition to the absorbing state ($Z = T$), the value of Z is simulated according to an exponential with rate parameter 1. Random censoring times were independently generated from uniform distributions $U[0, 3]$ and $U[0, 4]$ to introduce heavier or lighter censoring. To be more specific, in the case of the first scenario, the censoring proportion in the first gap time is 32% while for the second gap time, this percentage increases to 57%. The second model decreases these censoring levels to 24% and 47%, respectively. For each simulated scenario, we fixed the values of (s, t) pairs for four different points, corresponding to combinations of the percentiles of 20%, 40%, 60% and 80% of the exponential marginal distribution functions with rate parameter 1 to summarize the results of the transition probabilities. In each simulation, 1000 samples were generated with sample sizes of 50, 100, and 250. From these samples, we obtained the mean for all the generated data sets. As a measure of efficiency, we took the Mean Squared Error (MSE) but we also computed the standard deviations (SD) and the Bias for each point (s, t).

Tables 1 and 2 show the results of the LM and GG estimators for the transition probabilities $\widehat{p}_{12}(s, t)$, $\widehat{p}_{13}(s, t)$ and $\widehat{p}_{23}(s, t)$ from the simulation studies described previously. The parametric estimators, labeled as GG, were implemented by fitting a generalized gamma distribution to the multi-state model under the landmark approach through the `flexsurvreg` function of the `flexsurv` of the R package [16]. Results reveal that the accuracy of the estimators for all transition probabilities decreases as s and t grow due to the censoring effects that are higher at the right tail of the lifetime distribution. As would be expected, the SD decreases as the sample sizes increase, and with a decrease in the censoring percentage. From the results we can observe that the performance of the proposed parametric estimators is better than the nonparametric counterpart for cases with higher lag

times $t-s$, and consequently can be an alternative in case with small sample sizes when comparing with the LM estimator. This can be observed by the relative efficiency that is represented by the ratio between the corresponding MSEs of the nonparametric and parametric estimators. For the transition probabilities $\widehat{p}_{12}(s,t)$ and $\widehat{p}_{23}(s,t)$ the results are quite similar, revealing the better accuracy of the parametric estimator as the difference between times increases. In the case of $\widehat{p}_{13}(s,t)$, for almost all transition times the MSE ratio is above 1, confirming the advantages of the parametric estimator compared to the LM estimator. For completeness purposes, we show in Figures 5 and 6 the boxplots of the estimates of the transition probabilities based on the 1000 Monte Carlo replicates for the two sets of estimators, with different sample sizes. The boxplots shown in these figures reveal some results which agree with our findings reported in Tables 1 and 2. From these plots, it can be seen that all methods have small biases and confirm the lower variability of the parametric estimators in some cases. As expected, the lower variability of the parametric estimator comes at the cost of a small increase in the bias. This is more clear when the difference $t-s$ is small or for small sample sizes, and consequently, in these cases, the LM nonparametric estimator may be preferable.

4 Example of application

In this section, we analyse the results of the application of the proposed methods using data from a colon cancer cancer study. This is a study from a large clinical trial on Duke's stage III patients [24] that focuses on 929 patients who were followed after suffering curative surgery for colorectal cancer until death or censoring. Of the initial sample, 423 remained alive during the course of the follow-up period. 468 of the patients had a recurrence, and among these, 414 ended up dying. Finally, 38 individuals died without experiencing a recurrence.

Figure 7 reports, for each row, the estimated transition probabilities, $\widehat{p}_{11}(s,t)$, $\widehat{p}_{12}(s,t)$, $\widehat{p}_{13}(s,t)$ and $\widehat{p}_{23}(s,t)$, for fixed values of $s=365$ (left hand side) and $s=1095$ (right hand side). Each plot shows the estimated curves based on the parametric GG (black line) and the LM estimators (red line). Plots reported in the first row report the estimated transition probabilities for $p_{11}(s,t)$, for fixed values $s=365$ and $s=1095$ (days), along with time t. The estimated curves report the survival fraction over time among the individuals 'alive and without recurrence'. Plots in the second row allow for an inspection over time of the probability of being alive with recurrence for individuals who are disease free 1 year (left hand side) and 3 years (right hand side) after surgery. Since the recurrence state is transient, this curve first increases and then decreases. It is also evident that the probabilities for $s=1095$ are smaller than those for $s=365$. Finally, plots shown in the third and fourth rows report the estimated transition probabilities for $p_{13}(s,t)$ and $p_{23}(s,t)$. They report one minus the survival fraction along time, among the individuals 'alive and without recurrence' and those 'alive and with recurrence', respectively. Curves depicted in Figure 7 reveal that the parametric landmark estimators based on the generalized gamma distribution provide, in all cases, curves with the expected behavior, similar to those obtained from the nonparametric landmark estimators but with less variability.

5 Conclusion

In this article, the relative performance of the proposed parametric landmark estimators for the transition probabilities was investigated through simulations. Attained results suggest that the use of the generalized gamma distribution combined with the idea of subsampling can lead to competitive estimators that may outperform the original landmark estimators in some cases, providing estimators with less variability. It is worth mentioning that, in some cases, parametric estimation can introduce some bias in estimation while reducing the variance. The risk of introducing a large bias can be controlled in practice by comparing the obtained estimated curves with those obtained by the nonparametric estimator.

Acknowledgements: This research was financed by Portuguese Funds through FCT "Fundação para a Ciência e a Tecnologia", within the research grants PD/BD/142887/2018 and UID/BIA/04050/2019.

References:

[1] Andersen PK, Borgan O, Gill RD, Keiding N. Statistical Models Based on Counting Processes. Springer-Verlag, New York; 1993.

[2] Hougaard P. Analysis of multivariate survival data. Springer-Verlag. New York; 2000.

[3] Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time to event data. Statistical Methods in Medical Research; 2009; 18: 195–222.

[4] Meira-Machado L, Sestelo M. Estimation in the progressive illness-death model: a nonexhaustive review. Biometrical Journal; 2019; 61: 245–263.

[5] Andersen PK, Keiding N. Multi-state models for event history analysis. Statistical Methods in Medical Research; 2002; 11: 91–115.

[6] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine; 2007: 26, 2389–2430.

[7] Aalen OO, Johansen S. An empirical transition matrix for non homogeneous Markov and chains based on censored observations. Scandinavian Journal of Statistics; 1978; 5: 141–150.

[8] de Uña-Álvarez J, Meira-Machado L. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. Biometrics; 2015; 71: 141–150

[9] Putter H, Spitoni C. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. Statistical Methods in Medical Research; 2018; 27(7): 2081–2092.

[10] Aalen O, Borgan O, Gjessing H. Survival and Event History Analysis: A Process Point of View. Statistics. Springer-Verlag, New York. 2008.

[11] Benaglia T, Jackson C, Sharples L. Survival Extrapolation in the Presence of Cause Specific Hazards. Statistics in Medicine; 2015. 34 (5): 796–811.

[12] Datta S, Satten GA. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for non-Markov models. Statistics & Probability Letters; 2001; 55: 403–411.

[13] Borgan O. Aalen-Johansen Estimator. Encyclopedia of Biostatistics; 1988; 1: 5–10.

[14] Meira-Machado L,; de Uña-Álvarez J, Cadarso-Suárez C. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. Lifetime Data Analysis; 2006; 12: 325–344.

[15] Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Statistics in Medicine; 1992; 11: 1871–1879.

[16] Jackson C. flexsurv: A Platform for Parametric Survival Modeling in R. Journal of Statistical Software; 2016; 70 (8): 1–33.

[17] Cox C, Chu H, Schneider MF, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Statistics in Medicine; 2007; 26: 4352–4374.

[18] Kiche J, Oscar N, Orwa G. On Generalized Gamma Distribution and Its Application to Survival Data. International Journal of Statistics and Probability; 2019; 8 (5).

[19] Stacy AM. A Generalization of the Gamma Distribution. Annals of Mathematical Statistics; 1962; 33 (3): 1187–1192.

[20] Prentice RL. A Log Gamma Model and Its Maximum Likelihood Estimation. Biometrika; 1974; 61 (3): 539–544.

[21] Moreira A, de Uña-Álvarez J, Meira-Machado L. Presmoothing the Aalen-Johansen estimator in the illness-death model. Electronic Journal of Statistics; 2013; 7: 1491–1516.

[22] Araújo A, Meira-Machado L, Roca-Pardiñas J. TPmsm: Estimation of the transition probabilities in 3-state models. Journal of Statistical Software; 2014; 62(4), 1–29.

[23] Soutinho G, Meira-Machado L. Estimation of the Transition Probabilities in Multi-state Survival Data: New Developments and Practical Recommendations. WSEAS TRANSACTIONS on MATHEMATICS; 2020; 19, 353–366.

[24] Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, Veeder MH, Mailliard JA. Fluorouracil Plus Levamisole as Effective Adjuvant Therapy after Resection of Stage III. Colon Carcinoma: A Final Report, The Annals of Internal Medicine; 1995; 122 (5): 321–326.

Contribution of individual authors to the creation of a scientific article (Ghostwriting Policy)

Gustavo Soutinho and Luís Meira-Machado contributed equally to this work in terms of writing of the paper, methodological developments, and implementation of computer codes to obtain results from simulation studies and a real data set.

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)

Table 1: Bias and standard deviation (SD) for estimators of $p_{12}(s,t)$ and $p_{13}(s,t)$. The relative MSEs are also given. Scenario 1: illness-death model with correlated exponential gap times.

		$\widehat{p}_{12}^{\mathtt{LM}}(s,t)$	$\widehat{p}_{12}^{\mathtt{GG}}(s,t)$	Relative MSE	$\widehat{p}_{13}^{\mathtt{LM}}(s,t)$	$\widehat{p}_{13}^{\mathtt{GG}}(s,t)$	Relative MSE
		bias (SD)	bias (SD)	LM/GG	bias (SD)	bias (SD)	LM/GG
(s,t)=(0.2231,0.5108)							
n	C						
50	$U[0,4]$	0.0023 (0.0593)	0.0021 (0.0632)	0.8812	-0.0009 (0.0507)	-0.0150 (0.0437)	1.2049
	$U[0,3]$	0.0022 (0.0582)	0.0032 (NA)	0.7368	-0.0005 (0.0529)	-0.0105 (0.0470)	1.2055
100	$U[0,4]$	0.0013 (0.0404)	0.0009 (0.0407)	0.9884	0.0011 (0.0352)	-0.0122 (0.0287)	1.2781
	$U[0,3]$	0.0000 (0.0406)	0.0032 (0.0455)	0.7925	-0.0001 (0.0380)	-0.0095 (0.0330)	1.2191
250	$U[0,4]$	-0.0004 (0.0256)	-0.0016 (0.0244)	1.0939	-0.0009 (0.0223)	-0.0132 (0.0168)	1.0874
	$U[0,3]$	0.0012 (0.0272)	0.0049 (0.0300)	0.8014	0.0006 (0.0226)	-0.0088 (0.0194)	1.1235
(s,t)=(0.2231,1.6094)							
n	C						
50	$U[0,4]$	0.0002 (0.0823)	-0.0025 (0.0701)	1.3766	0.0009 (0.0940)	0.0164 (0.0838)	1.2110
	$U[0,3]$	0.0018 (0.0908)	-0.0055 (NA)	1.2650	-0.0012 (0.0981)	0.0112 (0,0881)	1.2197
100	$U[0,4]$	0.0010 (0.0583)	-0.0024 (0.0480)	1.4762	0.0001 (0,0659)	0.0171 (0.0582)	1.1783
	$U[0,3]$	-0.0015 (0.0660)	-0.0101 (0.0539)	1.4504	-0.0006 (0.0741)	0.0109 (0.0627)	1.3551
250	$U[0,4]$	-0.0013 (0.0385)	-0.0051 (0.0282)	1.8054	0.0010 (0.0408)	0.0184 (0.0348)	1.0729
	$U[0,3]$	-0.0003 (0.0406)	-0.0110 (0.0325)	1.3962	0.0007 (0.0427)	0.0138 (0.0362)	1.2130
(s,t)=(0.5108,0.9163)							
n	C						
50	$U[0,4]$	0.0005 (0.0837)	0.0102 (0.0987)	0.7124	0.0001 (0.0733)	-0.0139 (0.0700)	1.0559
	$U[0,3]$	0.0013 (0.0859)	0.0097 (NA)	0.7640	0.0033 (0.0775)	-0.0086 (0.0691)	1.2400
100	$U[0,4]$	0.0000 (0.0558)	0.0172 (0.0681)	0.6310	-0.0015 (0.0519)	-0.0123 (0.0483)	1.0840
	$U[0,3]$	0.0023 (0.0581)	0.0185 (0.0653)	0.7330	-0.0037 (0.0514)	-0.0056 (0.0516)	0.9844
250	$U[0,4]$	-0.0005 (0.0357)	0.02970.0484)	0.3967	-0.0003 (0.0334)	-0.0123 (0.0318)	0.9580
	$U[0,3]$	0.0003 (0.0386)	0.0264 (0.0478)	0.4994	-0.0003 (0.0332)	-0.0035 (0.0352)	0.8779
(s,t)=(0.5108,1.6094)							
n	C						
50	$U[0,4]$	0.0018 (0.1025)	0.0127 (0.1053)	0.9340	0.0019 (0.1103)	0.0156 (0.1061)	1.0591
	$U[0,3]$	-0.0001 (0,1142)	0.0119 (NA)	1.0281	-0.0020 (0.1201)	0.0085 (0.1063)	1.2683
100	$U[0,4]$	-0.0014 (0.0707)	0.0039 (0.0654)	1.1671	0.0031 (0.0743)	0.0179 (0.0691)	1.0846
	$U[0,3]$	0,0027 0,0800	0-0001 (0.0686)	1-3597	-0.0025 (0.0825)	0.0131 (0.0731)	1.2356
250	$U[0,4]$	0.0032 (0.0437)	0.0034 (0.0406)	1.1583	-0.0030 (0.0472)	0-0194 (0.0460)	0.8972
	$U[0,3]$	0.0002 (0.0477)	-0.0112 (0.0435)	1.1324	0.0009 (0.0508)	0.0188 (0.0455)	1.0635

Table 2: Bias and standard deviation (SD) for estimators of $p_{23}(s,t)$. The relative MSEs are also given. Scenario 1: illness-death model with correlated exponential gap times.

		$\widehat{p}_{23}^{\mathtt{LM}}(s,t)$	$\widehat{p}_{23}^{\mathtt{GG}}(s,t)$	Relative MSE	$\widehat{p}_{23}^{\mathtt{LM}}(s,t)$	$\widehat{p}_{23}^{\mathtt{GG}}(s,t)$	Relative MSE
		bias (SD)	bias (SD)	LM/GG	bias (SD)	bias (SD)	LM/GG
(s,t)=(0.2231,0.5108)					(s,t)=(0.5108,0.9163)		
n	C						
50	$U[0,4]$	-0.0085 (0.2378)	-0.0340 (0.2640)	0.7991	-0.0053 (0.1969)	-0.0093 (0.2238)	0.7734
	$U[0,3]$	-0.0130 (0.2509)	-0.0320 (0.2782)	0.8051	-0.0010 (0.1961)	0.0065 (0.2248)	0.7607
100	$U[0,4]$	0.0029 (0.1678)	-0.0022 (0.2080)	0.6507	-0.0034 (0.1343)	0.0053 (0.1468)	0.8367
	$U[0,3]$	-0.0114 (0.1653)	-0.0253 (0.1932)	0.7237	0.0013 (0.1363)	0.0053 (0.1418)	0.9223
250	$U[0,4]$	0.0020 (0.0976)	0.0018 (0.1091)	0.7997	-0.0019 (0.0829)	0.0177 (0.0813)	0.9926
	$U[0,3]$	0.0012 (0.1036)	0.0042 (0.1107)	0.8740	0.0015 (0.0835)	0.0199 (0.0812)	0.9967
(s,t)=(0.2231,1.6094)					(s,t)=(0.5108,1.6094)		
n	C						
50	$U[0,4]$	-0.0112 (0.1676)	0.0486 (0.1189)	1.7101	-0.0022 (0.1835)	0.0622 (0.1790)	0.9376
	$U[0,3]$	-0.0395 (0.1920)	0.0402 (0.1478)	1.6382	-0.0090 (0.1912)	0.0700 (0.1780)	1.0021
100	$U[0,4]$	-0.0064 (0.1190)	0.0328 (0.0878)	1.6177	0.0023 (0.1188)	0.0482 (0.1151)	0.9070
	$U[0,3]$	-0.0104 0.1290	0.0282 (0.0993)	1.5693	0.0085 (0.1284)	0.0544 (0.1208)	0.9426
250	$U[0,4]$	0.0025 (0.0723)	0.0016 (0.0578)	1.5665	-0.0010 (0.0738)	0.0136 (0.0619)	1.3597
	$U[0,3]$	-0-0038 (0.0790)	-0.0008 (0.0626)	1.5985	0.0016 (0.0798)	0.0094 (0.0713)	1.2329

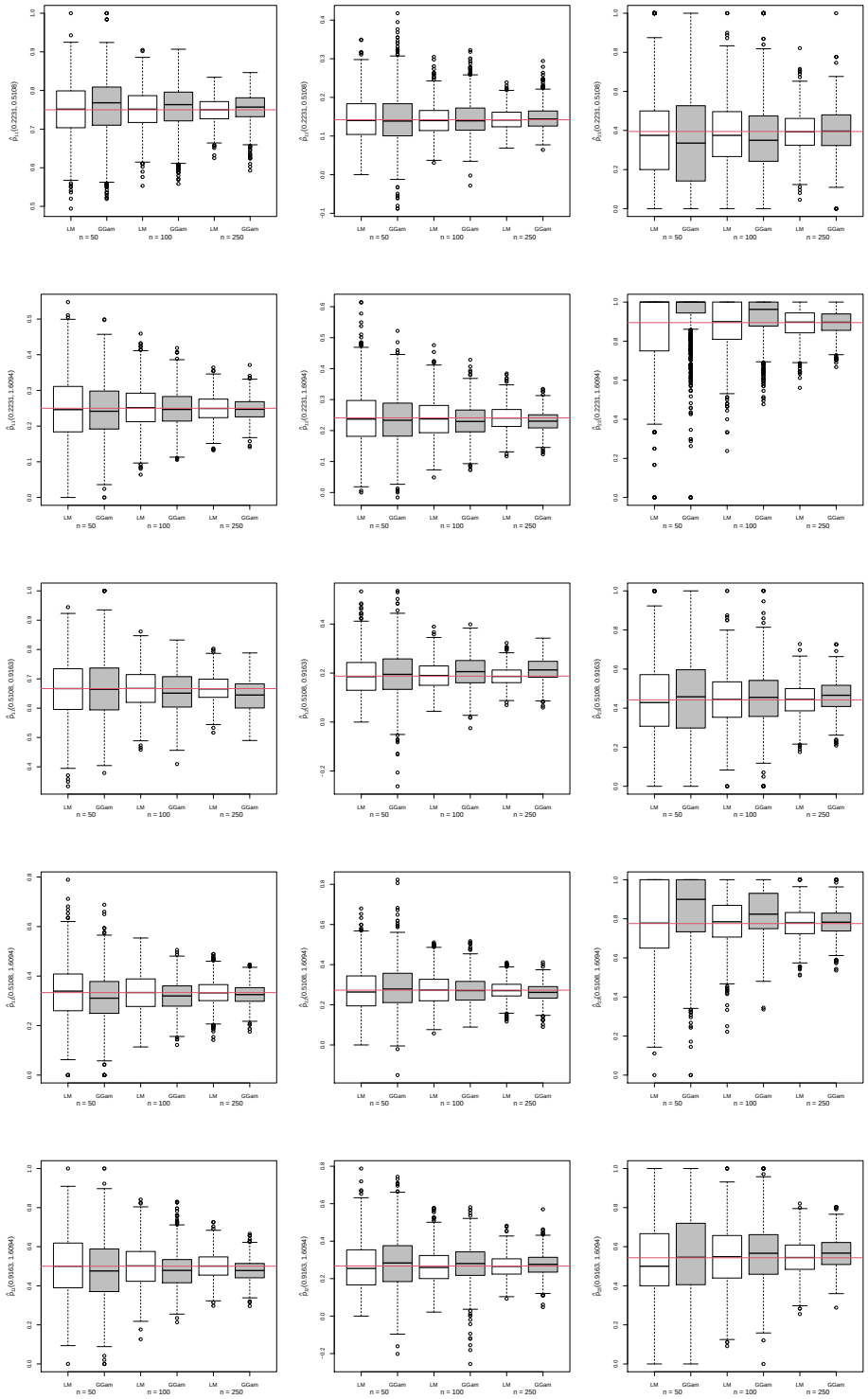

Figure 5: Boxplot of the $M = 1000$ estimates of the transition probabilities. Horizontal solid red line correspond to the true value of the transition probability. Censoring times uniformly distributed between 0 and 3.

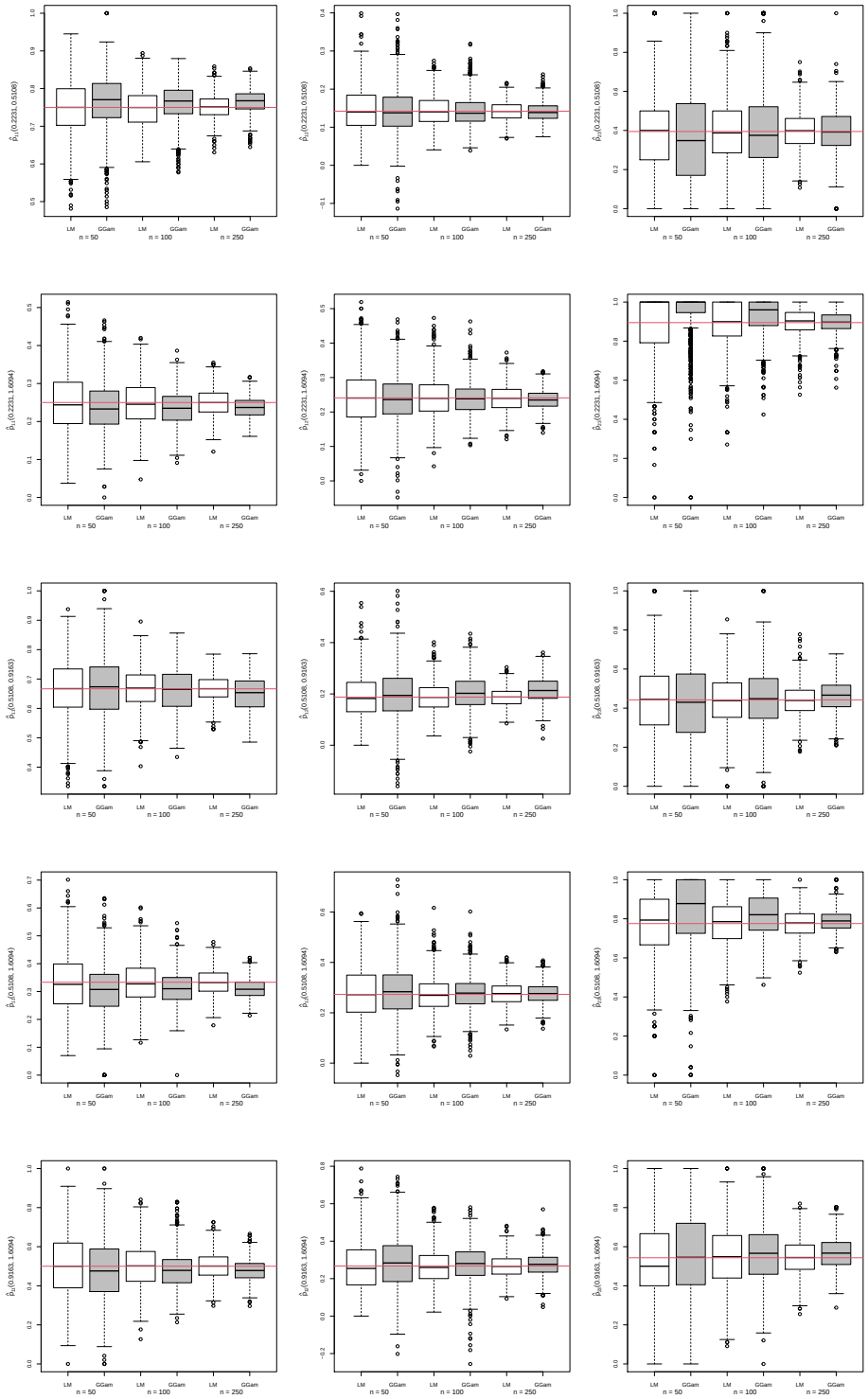

Figure 6: Boxplot of the $M = 1000$ estimates of the transition probabilities. Horizontal solid red line correspond to the true value of the transition probability. Censoring times uniformly distributed between 0 and 4.

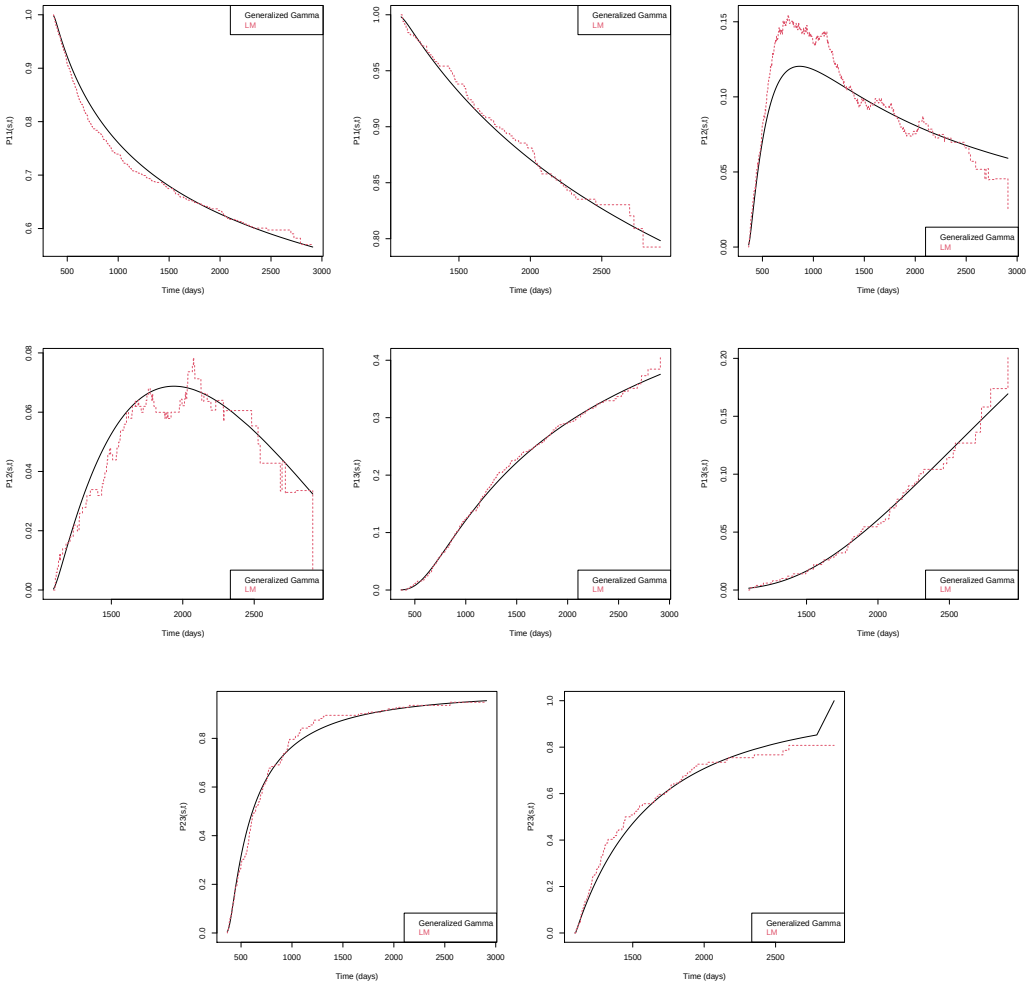

Figure 7: Estimated transition probabilities for $s = 365$ and $s = 1095$. ColonIDM data.