

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Juliana Freitas de Mello e Silva

*Modelo Exponencial por Partes para Dados de Sobrevida
com Longa Duração*

Piecewise Exponential Model for Long Term Survival Data

Belo Horizonte

2016

Juliana Freitas de Mello e Silva

*Modelo Exponencial por Partes para Dados de Sobrevida
com Longa Duração*

Piecewise Exponential Model for Long Term Survival Data

Dissertação apresentada ao Curso de Estatística da
UFMG, como requisito para a obtenção do grau de
MESTRE em Estatística.

Orientador: Fábio Nogueira Demarqui

Doutor em Estatística

Co-orientador: Thiago Rezende dos Santos

Doutor em Estatística

Belo Horizonte

2016

Freitas de Mello e Silva, Juliana

Modelo Exponencial por Partes para Dados de Sobrevivência com
Longa Duração / Juliana Freitas de Mello e Silva - 2016

73.p

1.Estatística. I.Título.

CDU 519.5

Juliana Freitas de Mello e Silva

*Modelo Exponencial por Partes para Dados de Sobrevida
com Longa Duração*

Piecewise Exponential Model for Long Term Survival Data

Dissertação apresentada ao Curso de Estatística da
UFMG, como requisito para a obtenção do grau de
MESTRE em Estatística.

Aprovado em 22/02/2016.

BANCA EXAMINADORA

Fábio Nogueira Demarqui

Doutor em Estatística

Thiago Rezende dos Santos

Doutor em Estatística

Wagner Barreto de Souza

Doutor em Estatística

Leonardo Soares Bastos

Doutor em Estatística

À minha família.

Resumo

O Modelo Exponencial por Partes (MEP) é um modelo bastante utilizado, principalmente em análise de sobrevivência. Ao utilizar esse modelo, uma partição do eixo do tempo em um número finito de intervalos é estabelecida e, em seguida, uma taxa de falha constante é considerada para cada um dos intervalos. Portanto, o MEP aproxima uma função contínua, a saber a taxa de falha, através de seguimentos de reta. Por essa razão, o MEP é um modelo bastante flexível e, embora este seja um modelo paramétrico, é frequentemente considerado como não paramétrico. O presente trabalho propõe uma abordagem bayesiana dinâmica que permite a obtenção de uma distribuição suavizada exata para os parâmetros representando a taxa de falha. Além disso a partição do eixo do tempo (e, conseqüentemente, o número de intervalos) será considerada como uma quantidade desconhecida a ser estimada. Toda a abordagem proposta será utilizada para modelar a fração de cura em uma população, o que ocorre quando uma parte dos indivíduos em um estudo é considerada curada e, portanto, nunca experimentará o evento de interesse. Para que seja possível uma comparação, o caso com grade fixa também será considerado. Por fim, será mostrada uma aplicação a fim de ilustrar os conceitos apresentados.

Palavras-chaves: análise de sobrevivência, modelo de riscos proporcionais, abordagem dinâmica, modelo de fração de cura, modelo exponencial por partes.

Abstract

The Piecewise Exponential Model (PEM) is a very utilized model, mainly in survival analysis. When using this model, one considers a partition of the time axis into a finite number of intervals and, after that, a constant failure rate is considered to each interval. Therefore the PEM approximates a continuous function, the failure rate, through line segments. For this reason, the PEM is a very flexible model and, although it is a parametric model, it is often considered as non parametric one. The present work proposes a Bayesian dynamic approach that allows one to obtain the exact smoothed distribution for the parameters representing the failure rate. Moreover, the partition of the time grid (and, consequently, the number of intervals), will be considered as an unknown quantity to be estimated. This entire approach will be used to model the cure fraction in a population, which occurs when a part of the individuals in a study is considered cured and, therefore, will never experience the event of interest. For comparison purposes, the fixed time grid will also be considered. Lastly, in order to illustrate this approach, an application will be shown.

Keywords: survival analysis, proportional hazards models, dynamic approach, cure fraction model, piecewise exponential model.

Agradecimentos

Começo meus agradecimentos com a frase “Yeah, I’m a lucky man, to count on both hands the ones I love. Some folks they’ve got one, yeah, others, they’ve got none.”. Ela me lembra do quanto devemos ser gratos sermos por quem somos, por quem nos cerca e pelo que possuímos, seja muito ou pouco.

Gostaria de agradecer à minha família. Principalmente à minha mama Isabel, por ter me apoiado nessa parte tão importante na minha vida. Pelo incentivo de sair de casa e (v)ir para BH, por aturar meus chororôs de nervosismo ao telefone (e pessoalmente :D). Agradeço também à minha tia Cássia pela força a aprovação que são tão importantes pra mim. Agradeço aos meus irmãos Filipe, Mateus e Tiago e ao meu pai Geraldo.

À minha amiga de tanto tempo Isabella por estar sempre perto de mim, mesmo de longe. À Juliana pela amizade e por compartilhar um pouco das minhas loucuras. Ao amigo Lucas por ser capaz de sempre colocar um sorriso no meu rosto, obrigada :). À Jordana por se importar comigo e me ouvir.

Agradeço aos amigos que fiz durante a graduação na UFF. Ao Bruno, pela amizade, pelo apoio, pelas conversas e por estar sempre presente. À Fernanda eu agradeço por todas as visitas, sejam elas em Minas ou no Rio (quero mais visitas!). À Carol e a Keilane pela amizade que continua, mesmo nos encontrando tão pouco. Vocês são muito importantes pra mim!

Agradeço imensamente aos membros da Liga Externa: à querida Caroline Ponce (e sua família!), ao Luiz Fernando e ao Rafael Erbisti. Realmente sou agradecida pela paciência e preocupação que tiveram comigo, pela troca de conhecimento, por toda a ajuda, incentivo e apoio. Ah, pelas pizzas de toda santa terça também! Foi muito bom e fundamental ter vocês junto comigo.

Aos amigos que fiz nesse mestrado. À Fernanda por estar tanto tempo do meu lado, pelas conversas, pelas histórias, pelo incentivo e pela companhia. Ao Estevão por TODAS as conversas/discussões que me mostravam novos pontos de vista e me faziam crescer. À Gabi e à Bárbara pelo apoio, pela doçura e pela companhia durante esse

mestrado. Ao Douglas pela amizade, pelas risadas e por toda a ajuda fornecida. Ao Edson, Guilherme, Milton, Uriel e demais colegas pela companhia e pelo conhecimento. Agradeço também à Rafaella, ao Cristiano e ao Zé Luiz pelas conversas e pela companhia. À Silvana pela gentileza, simpatia e troca de conhecimento. Com certeza esse mestrado foi uma festa com vocês.

Agradeço à Rachel por ter me ensinado muito sobre a vida, pela companhia, pelas risadas e pelas discussões filosóficas antes de dormir.

Ao Paulo, eu agradeço por toda ajuda! Por toda a troca de conhecimento que foi muito boa e muito importante, tanto na parte dos conceitos quanto na parte computacional. Agradeço também o incentivo e otimismo.

Agradeço à professora Alexandra Schmidt por mostrar conceitos de forma tão simples e com tanto entusiasmo. Por mostrar também a importância de conceitos sociais, algo que normalmente é esquecido no dia a dia. Agradeço também pela conversa e apoio.

Sou grata ao Fábio Demarqui e ao Thiago Rezende por terem possibilitado o conhecimento de conceitos tão interessantes através da oportunidade de trabalhar nesse projeto. Pelas reuniões e pelo incentivo.

Agradeço aos professores Valetin Sisko e Leonardo Bastos pela solicitude ao me fornecerem a carta de recomendação. Ao professor Wagner Barreto pelo excelente curso de Probabilidade, consegui aprender conceitos de uma área que considero tão complicada. Agradeço novamente ao Leonardo e ao Wagner por aceitarem o convite e fazerem parte da banca.

Agradeço também, de forma geral, à todas as pessoas que passaram pela minha vida, que torceram por mim e que, de alguma forma, fizeram com que eu chegasse até aqui.

Por último e realmente não menos importante, agradeço à CAPES pelo auxílio financeiro que, sem o qual, eu não teria cursado esse mestrado.

The future is paved with better days.

Eddie Vedder

Sumário

Lista de Figuras	8
Lista de Tabelas	9
1 Introduction	10
1.1 Purposes	13
2 Basic Concepts in Survival Analysis	14
2.1 Proportional Hazards Model	16
2.2 Cure Fraction Models	17
3 The Piecewise Exponential Model	25
3.1 Random Time Grid	32
4 Applications	41
4.1 Brain Cancer Data	42
4.2 Melanoma Data	48
5 Conclusions and Future Works	58
A Appendix	60

Lista de Figuras

2.1	Comparison of the follow-up time and the median survival time for the uncured individuals.	18
3.1	Illustration of the quantity t_j	26
3.2	Illustration of the intervals' grouping scheme.	34
4.1	Comparison of the failure rate estimate varying the number of intervals. . .	43
4.2	Boxplot of the posterior discount factor sample for fixed and random time grid varying the (maximum) number of intervals.	44
4.3	Model comparison measures for the fixed time grid x random time grid according to different (maximum) number of intervals.	45
4.4	Histograms of the number of intervals varying the maximum number of intervals.	46
4.5	Estimated survival function.	47
4.6	Comparison of the failure rate estimates using fixed and random time grid and also varying the (maximum) number of intervals.	50
4.7	Model comparison measures for the fixed time grid x random time grid according to different (maximum) number of intervals.	51
4.8	Comparison of the probability of cure estimated varying the number of intervals.	52
4.9	Histogram of the number of intervals varying the maximum number of intervals.	53
4.10	Comparison of the discount factor estimates varying the number of intervals.	54
4.11	Comparison of the probability of cure separated by the covariates age and sex.	55
4.12	Estimated survival functions.	56

Lista de Tabelas

4.1	Comparison of fixed time grid x random time grid according to different (maximum) number of intervals.	45
4.2	Summary of the number of intervals.	46
4.3	LPML and WAIC results.	49
4.4	Estimates of the intercept (ψ_1) for the case with random time grid varying the maximum number of intervals.	51
4.5	Summary of the number of intervals.	52
4.6	Estimates of the coefficients associated to the non-cured individuals and the cure fraction.	55

1 Introduction

Survival analysis is an area of statistics utilized when the intent is to study the time until the occurrence of an event of interest. Several studies involving survival analysis are concentrated in the medical area, but there are also important applications on engineering, economy, quality control, among others.

Survival data have intrinsic characteristics such as asymmetry and incomplete observations, called *censure*. Therefore, specific methods are required. One of the important quantities related to this area is the hazard function, or failure rate. This function provides the risk that the event of interest has to happen and it brings along with it a challenge: when modeling parametrically, the most utilized models impose a particular form, or few particular forms, for this function. For example, by choosing the Weibull or Gamma model it is possible to obtain an increasing, a decreasing or a constant hazard function. Other distributions, such as the Birnbaum-Saunders (Birnbaum and Saunders, 1969) and the Generalized Weibull are richer in form, however, the challenge still remains. It arises when the chosen model and the shapes that it carries are not the most suitable one to fit the data, thus this function would not be well represented.

A fine alternative to model survival data is the Piecewise Exponential Model (PEM). Basically, to define this model one has to partition the time axis into b intervals and to assume a constant failure rate in each interval. By using this model one is approximating the failure rate by line segments. For this reason, the hazard function of the PEM does not have a pre-determined form, providing great flexibility in the survival data modeling. Based on this characteristic, the PEM is often regarded as a nonparametric model, although in fact, it is a parametric one.

Thereby, this model has been standing out in the literature in the recent years and numerous extensions have been proposed. Those extensions concern some important features of the model, such as: how to partition the time axis and, consequently, the number of intervals; how to estimate the hazard function; the inclusion of covariates (which may be time dependent or not); cure fraction models and so on.

The number of intervals must be chosen with caution: if it is too large there will be few data in each interval, leading to poor and/or unstable estimates; on the other hand, if it is too small, the true shape of the hazard function may not be achieved. A good number must balance the quantity of data in each interval, providing a good estimation for both hazard and survival functions (Demarqui, 2010). Alternatively, one can estimate the partition of the time axis and, consequently, the number of intervals. In most works found in the literature the time grid is chosen arbitrarily, fixing the number of intervals. Some of these works are: Gamerman (1991), Gamerman (1994), Ibrahim et al. (2001a), Yin and Ibrahim (2005), de Castro et al. (2009) and de Castro et al. (2015). However, there are also those who estimate it: Arjas and Gasbarra (1994), McKeague and Tighiouart (2000), Kim et al. (2007) and Demarqui (2010) estimate it through different approaches. Kim et al., for example, make use of Reversible Jump while on Demarqui's work, the Product Partition Model (PPM) is used.

The estimation of the hazard function can be done by using different approaches within the Bayesian context: one can assume that the components of this function are independent; a dynamic model can be used, that is, it is possible to carry information along the intervals; historical data can also be used to construct the prior distribution (see Ibrahim et al. (2001b)).

Yin and Ibrahim (2005), for example, consider the components of the hazard function as independent *a priori*. Gamerman (1991) and Gamerman (1994) considers a dynamic approach to estimate the hazard function, which relates each component of the hazard function through an evolution equation. The difference between these two works relies on the evolution equation: in the first work, the author includes time-dependent covariates, making the hazard function to be a function that depends only on the (time dependent) coefficients and covariates; in the second one, however, there are no covariates, thus the hazard function of an interval depends only on the hazard function of the previous one. de Castro et al. (2009) also considers the components of the hazard function as dependent. In Demarqui (2010) there is a comparison of different types of prior distributions for the hazard function: independent *Gamma* prior, Jeffrey's prior, prior within the dynamic approach and structural prior. In all works that used the dynamic model, an online and an *approximated* smoothed distribution were obtained.

Cure fraction models were developed to adapt the cases of long-term survivors, that is, those cases in which there are individuals who will never experience the event of

interest. By using such models it is possible to obtain information about the factors that influence the cured individuals as well as those factors related to the non-cured ones, separately; and also the probability of cure. Those information are very important to patients who suffer from cancer, for example.

There are two approaches concerning the cure fraction: the Mixture and the Promotion Time models. First, Boag (1949) and Berkson and Gage (1952) proposed the mixture cure rate model. In a brief way, this approach splits the population into two sub-populations: one is composed by the cured individuals and the other, by the non-cured ones. The Promotion Time Model, in turn, was introduced in 1996, by Yakovlev and Tsodikov based on a attractive biological motivation. Chen et al. (1999) extended this model for the Bayesian context. Ibrahim et al. (2001b) show an approach that allows the link between these two models.

There are several extensions and applications involving cure rate models: in Ibrahim et al. (2001a) the authors introduce a parameter to control the right tail of the survival curve, in a way that it becomes possible to control the degree of parametricity in the beginning, in the middle and, most importantly at the end of the survival distribution. Banerjee and Carlin (2004) use cure fraction model in the context of spatial analysis applied to a smoking cessation study. Yin and Ibrahim (2005) propose a cure fraction model that allows one to obtain zero and non zero cure fraction estimates, that is, there is no need to assume a cure fraction, the proposed model engages both cases. In Basu and Tiwari (2010) there is an extension involving competing risks and an interesting application involving breast cancer patients. In turn, Cucchetti et al. (2015) applied cure fraction model to study patients that suffered from colorectal liver metastases.

Nevertheless, in most works found in the literature (Kim et al. (2007), de Castro et al. (2009), Demarqui et al. (2014) and others) the information from the covariates is used only to explain the cure fraction. It would be interesting to observe how these covariates influence the non-cured individuals as well.

In this present work, the PEM will be used to model the data along with the dynamic approach, in a way that an *exact* smoothed distribution will be obtained for the hazard function, unlike the other works aforementioned. Regarding the time grid both cases will be considered: fixed and random (via PPM). Moreover, long term survivors will

be considered and the information from covariates will be used to explain both cured and non-cured individuals.

This work is organized as follows. Chapter 2 explicits the basic theory in which this work is based on. The concepts showed in this Chapter will be essential for understanding this work. In Chapter 3 the Piecewise Exponential model is explored and discussed. In turn, in Chapter 4 two applications are presented with the aim to illustrate the concepts presented. At last, Chapter 5 concerns the conclusions obtained as well as the future works.

1.1 Purposes

The aim of this study is to consider the Piecewise Exponential Model in the Bayesian dynamic approach. In the present work the parameters representing the hazard function will be correlated so that the information of the previous interval can be used to estimated the actual one. By doing this, a quantity to control the passage of information is introduced. This quantity is called discount factor and it will be estimated, differently from most works in the literature. The evolution equation used in this work allows the achievement of an exact smoothed distribution, this is also commonly obtained in an estimated way.

The time grid of the PEM will be estimated via Product Partition Model and the results obtained will be compared to the fixed time grid case. In turn, the whole modeling procedure obtained will be used to study cure fraction models. These studies are standing out in the literature, since new and more effective treatments are emerging.

The ultimate purpose is to infer about the pros and cons of the proposed model, discovering which are the cases that this model is more appropriate.

2 Basic Concepts in Survival Analysis

In this chapter, basic concepts and properties used in survival analysis will be introduced and discussed. These concepts will be essential for understanding this present work.

Survival analysis can be used when the interest is to study the time until the occurrence of a certain event. Let T be the non negative random variable representing the observed time. This time may be a failure time or a censored time. A failure time occurs when the event is fully observed, whereas a censored time takes place when the time to event is, for some reason, not fully observed due to, for instance, by the end of the study, the lost of follow-up, death of the patient for some other reason than that particular one of interest, among others (Colosimo and Giolo, 2006; Carvalho et al., 2011). Another typical characteristic found in survival data is asymmetry, which makes impracticable the usage of common methods, generally involving the Normal distribution.

There are three types of censoring, namely: left censoring, right censoring and interval censoring. Left censoring occurs when the event of interest has happened before the time to event is observed, for example, in a study where the event is the first time the individual has smoked a cigarette, the individual may not remember when it was, but it is known for sure that, if it happened, it happened before the interview; right censoring occurs when the actual time to event is known to be greater than the observed time; in turn, interval censoring occurs when it is known that the time to event has occurred in a interval, for example, when the study considers seropositive patients and the interest is to evaluate when the progression to AIDS will happen, it will be known that this time will be between two exams, but the exact time will be unknown.

The most frequent case in practice is the right censoring scheme, which will be explored in this work along with a non-informative mechanism, that is, the censorship mechanism is not related to the time to event. A deeper explanation about types of censoring can be found in Lawless (2002). Therefore, the observed time in this case is the minimum between the time to failure and the time to censorship.

An indicator variable, δ_i , will be used to represent whether the $i - th$ time is a failure or a censored time, that is:

$$\delta_i = \begin{cases} 1, & \text{if the failure occurred for the } i - th \text{ individual,} \\ 0, & \text{if the } i - th \text{ individual is right censored.} \end{cases}$$

Therefore, for each individual the information will be the pair (t_i, δ_i) , where i represents the $i - th$ individual. If there is information from covariates, the information will be represented by $(t_i, \delta_i, \mathbf{x}_i)$, where \mathbf{x}_i is the covariates vector associated to the $i - th$ individual.

Let T be a continuous non-negative random variable whose probability density function (p.d.f.) is $f(t)$. The survival function is defined as:

$$S(t) = P(T > t) = 1 - F(t), t > 0. \quad (2.1)$$

This function has the following well-known properties:

1. $S(0) = 1$;
2. $\lim_{t \rightarrow \infty} S(t) = 0$;
3. $S(t)$ is a decreasing function in t .

The first property means that at time 0, all the individuals have not suffered the event of interest (for example, if the event is death from breast cancer, all individuals have not died from this disease); the second property means that as $t \rightarrow \infty$, all individuals will eventually suffer the event of interest.

The hazard function is defined as the instantaneous rate of failure at time t . Its expression is given by:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}, t > 0. \quad (2.2)$$

Another important function is the cumulative hazard function, which is given by

$$H(t) = \int_0^t h(u) du, t > 0. \quad (2.3)$$

These functions satisfy the following relationships:

$$S(t) = \exp\left(-\int_0^t h(u)du\right),$$

$$f(t) = -\frac{d}{dt}S(t),$$

$$h(t) = \frac{d}{dt}\log(S(t)).$$

Also, note that:

$$\begin{aligned} f(t) &= \frac{d}{dt}F(t) = \frac{d}{dt}[1 - S(t)] = -\frac{d}{dt}S(t) = -\frac{d}{dt}\exp\left(-\int_0^t h(u)du\right) \\ &= -\left[\exp\left(-\int_0^t h(u)du\right)(-h(t))\right] = S(t)h(t). \end{aligned} \quad (2.4)$$

That is, $f(t) = S(t)h(t)$ and then $h(t) = \frac{f(t)}{S(t)}$.

2.1 Proportional Hazards Model

One may think of including covariates into the model. In this case, a very important model is the proportional hazards (PH) model. This model was proposed by Cox (1972) and it includes the information from covariates through the hazard function, in the following way:

$$h(t|\mathbf{x}) = h_0(t) \exp(\mathbf{x}\boldsymbol{\beta}), \quad (2.5)$$

where $h_0(t)$ is the baseline hazard function, \mathbf{x} is the covariates vector and $\boldsymbol{\beta}$ is the vector of regression coefficients.

This model has the property of proportional hazards, that is, the hazard ratio of two individuals does not depend on time. This property can be seen through Equation (2.6), where \mathbf{x}_1 and \mathbf{x}_2 are the covariates vector of two different individuals.

$$\frac{h(t|\mathbf{x}_1)}{h(t|\mathbf{x}_2)} = \frac{h_0(t) \exp(\mathbf{x}_1\boldsymbol{\beta})}{h_0(t) \exp(\mathbf{x}_2\boldsymbol{\beta})} = \frac{\exp(\mathbf{x}_1\boldsymbol{\beta})}{\exp(\mathbf{x}_2\boldsymbol{\beta})}. \quad (2.6)$$

The baseline hazard can be modeled non-parametrically, as first proposed by Cox (1972), or parametrically. On the first case, the partial likelihood method is adopted as the likelihood function and used to estimate the coefficients (see more in Carvalho et al. (2011) and Klein and Moeschberger (2003)). On the other one, the likelihood function is constructed in the following manner: in the case of right censoring and a non-informative

mechanism, it is given by the p.d.f. for those individuals whose time to event is completely observed and the survival function for those whose time to event is (right) censored. That is:

$$L(\Phi; D) = \prod_{i=1}^n f(t_i|\Phi)^{\delta_i} (S(t_i|\Phi))^{1-\delta_i} = \prod_{i=1}^n h(t_i|\Phi)^{\delta_i} S(t_i|\Phi), \quad (2.7)$$

where Φ is the vector of parameters to be estimated which includes the vector of coefficients. D represents the data available, in this case $D = (t_i, \delta_i, \mathbf{x}_i)$, $i = 1, 2, \dots, n$; n is the number of individuals and, lastly, t_i represents the time to event of the i -th individual.

2.2 Cure Fraction Models

Over the past decades, with the advance of medicine, patients' survival is being improved and it implies directly on the probability of survival, usually raising it. For this reason, a considerable part of patients is being cured. It is important to highlight that the concept of "cure" is not strictly medical; in fact, if an individual is considered cured it means that the event will never happen to this specific individual. This characteristic violates the second property of the survival function, that is, the survival function does not go to 0 as $t \rightarrow \infty$, it goes to the proportion of healed individuals, which will be represented by $\pi \in [0, 1]$.

Cure fraction models were developed to adapt these situations. In the literature there are several articles that analyze data with a cure fraction, such as Farewell (1982), Farewell (1986), Ibrahim et al. (2001a), Kim et al. (2007) and others.

The challenge, at this point, is to separate the truly cured individuals from those who have not suffered the event yet, due to the duration of the follow-up, for example. According to Yu et al. (2004), the efficiency of the estimate of the cure rate depends, among other factors, on the follow-up time. Cases in which the follow-up time is relatively greater than the median of the survival time for the uncured individuals are the cases in which the cure rate presents better estimates. This is due to the confusion between the truly cured individuals with those who just have not suffered the event of interest yet (but they would, if the follow-up time was long enough).

Figure 2.1 allows a visualization of the discussion presented. In that figure, there is the Kaplan Meier estimate of the survival function based on the E1673 dataset,

described in Ibrahim et al. (2001b), in which the event of interest is death from melanoma. The dotted red line represents the median survival time for the non-cured individuals. Note that the follow-up time is considerably larger than the median survival time.

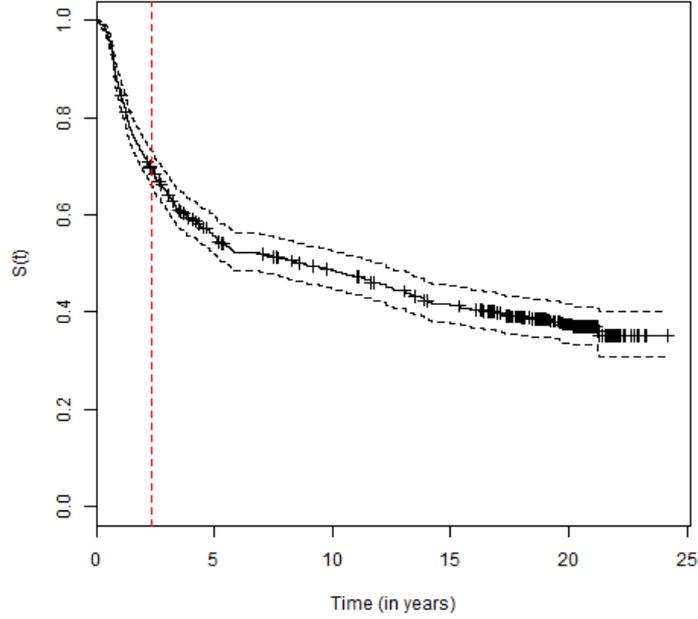


Figure 2.1: Comparison of the follow-up time and the median survival time for the uncured individuals.

Moreover, Lambert (2007) points out that when using cure fraction models one is assuming that there is a cure, nevertheless it may not be medically correct. For example, studies in which the interest is to evaluate the time to death from a certain type of cancer but it is only recorded if the individuals had died or not, the cause of death is unknown; in this case one would be assuming that exists cure from death, which is senseless.

Therefore, to introduce this model, consider a non-negative function $f^*(t)$ such that $\int_0^\infty f^*(t)dt = 1 - \pi \leq 1$ and, given that, the adapted survival function, called improper survival function will be (Rodrigues et al., 2008):

$$S_{pop}(t) = \pi + \int_t^\infty f^*(u)du \quad (2.8)$$

This function will have the following properties:

1. If $\pi = 0$, then $S_{pop}(t) = S(t)$,

2. $S_{pop}(0) = 1$;
3. $S_{pop}(t)$ is a decreasing function in t ;
4. $\lim_{t \rightarrow \infty} S_{pop}(t) = \pi$.

The first property refers to the case when there's no cure fraction, thus the analysis will be the same as the one presented before; therefore, $S(t)$ is a genuine survival function. The meaning of the second and third properties is analogue to the proper survival function and, the fourth one means that when $t \rightarrow \infty$, the survival function goes to the proportion of the cured individuals.

The most important cure fraction models found in the literature are the mixture and the promotion time models. These models carries advantages and disadvantages: the mixture model is quite logical and the promotion time model has an interesting biological motivation that enriches the interpretation in a study. However, as pointed out by Rodrigues et al. (2008), the mixture model allows only one causing factor of the event of interest whereas the promotion time model allows more than one. Besides that, the assumption of proportional hazards may no longer be preserved, mainly for the mixture model (Ibrahim et al., 2001b; Rodrigues et al., 2009). According to Ibrahim et al. (2001b), this assumption is no longer attained for the mixture model when the probability of cure is modeled through a binomial regression.

Mixture Model

The mixture model was proposed by Boag (1949) and Berkson and Gage (1952) and it is given by:

$$S_{pop}(t) = \pi + (1 - \pi)S(t). \quad (2.9)$$

The idea of this model is to include the individuals through two components: one representing the cured individuals and the other representing the non-cured ones along with its survival. In this way π represents the proportion of cured individuals and, consequently, $1 - \pi$ represents the non cured ones; $S(t)$ is the usual (and proper) survival function for the non-cured individuals, that is, $S(t) = \int_t^\infty \frac{f^*(u)}{1 - \pi} du$.

Consequently, the populational hazard function and the ‘‘p.d.f.’’ will be given by:

$$f_{pop}(t) = -\frac{d}{dt}S_{pop}(t) = (1 - \pi)f(t), \quad (2.10)$$

$$h_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)}. \quad (2.11)$$

Note that Equation (2.10) is not a proper p.d.f. indeed, because $\int_0^\infty f_{pop}(t) = (1 - \pi) \int_0^\infty f(t) = 1 - \pi \leq 1$.

So, substituting $S_{pop}(t)$ and $f_{pop}(t)$ in Equation (2.7), the likelihood function for n individuals will be given by:

$$\begin{aligned} L(\Phi, \psi; D) &= \prod_{i=1}^n [(1 - \pi_i)f(t_i|\Phi)]^{\delta_i} [\pi_i + (1 - \pi_i)S(t_i|\Phi)]^{1-\delta_i} \\ &= \prod_{i=1}^n [(1 - \pi_i)S(t_i|\Phi)h(t_i|\Phi)]^{\delta_i} [\pi_i + (1 - \pi_i)S(t_i|\Phi)]^{1-\delta_i}, \end{aligned} \quad (2.12)$$

where Φ is the vector including the parameters that indexes f and the coefficients of the non-cured individuals. In turn, $\pi_i = g(\mathbf{z}_i\psi)$ for some link function $g(\cdot)$, this link function may be the *logit*, for example. The covariates associated to the cure fraction can be different from the ones used to model the non-cured individuals. Therefore to model the cure fraction, let \mathbf{z}_i represent the vector of covariates of the i -th individual and, consequently, ψ is the coefficients' vector, including the intercept. Moreover, in this case, $D = (t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$, for $i = 1, 2, \dots, n$.

It is noteworthy, in the Bayesian context, that the prior distribution for ψ , the vector of coefficients associated to the cure fraction, must be chosen carefully as improper priors may not lead to a proper posterior distributions (Ibrahim et al., 2001b). Moreover, few issues involving this model were found both in the literature and practice, such as convergence problems when using large variance for the cure fraction coefficients prior (Banerjee and Carlin, 2004) and identifiability problems (Klein et al., 2014).

A deep study regarding the identifiability of cure fraction models was presented by Li et al. (2001). These authors proved that the mixture model *is* identifiable in the case of the present study: that case where the model of the non-cured fraction is fully parametric. Nevertheless, they also state that this specific case represents a case of ‘‘near non-identifiability’’, such characteristic shows itself in the form of numerical issues and/or flat likelihood. Works such as Yu et al. (2004), also related difficulties when using this

model. Taylor (1995) and Peng (2003) describe solutions to this problem: in a very brief explanation these authors force the improper survival function to become a proper one.

Another discussion to justify the issues attached to the usage of the mixture model can be found in Klein et al. (2014). They point out the possible confusion obtained when using the same covariates to model the cure fraction and the non-cured individuals.

Therefore, taking these points into account, the focus will rely only on the promotion time model.

Promotion Time Model

The promotion time model was developed by Yakovlev and Tsodikov (1996) and its Bayesian version was proposed by Chen et al. (1999). This model has an interesting and appealing biological motivation: consider a scenario in which an individual that underwent a treatment due to a certain type of cancer. After this procedure there may remain some cancer cells that may become active again and develop a new tumor, in other words, the individual may relapse. Starting from this scenario, consider N as the number of competent cells, that is, those cells that can become a tumor, and $Z_l, l = 1, 2, \dots, N$ the time until the l -th competent cell will become active, this time is also called the promotion time. One characteristic of this model is that the number of competent cells is a latent random variable and, given this, $Z_l, l = 1, 2, \dots$ are considered to be independent and identically distributed (i. i. d.), with a cumulative distribution function $F(t)$, which does not depend on N . This distribution can be the Exponential or the Weibull, for example; in this present work it will be the Piecewise Exponential distribution.

The time to relapse is the time until the first competent cell becomes active, that is, $T = \min \{Z_l, 0 \leq l \leq N\}$, with $P(Z_0 = \infty) = 1$. In this way, if there is no competent cell, the subject is considered cured and therefore, the time until a competent cell becomes active is, certainly, infinite.

It is worth mentioning that the biological motivation does not exclude cases in which the event of interest is any other than relapse. According to Ibrahim et al. (2001b), this model can be used to every situation in which it is considered to exist a cure fraction and there exist N competing risks. Competing risks are referred to cases in which it is known that the individuals are exposed to several outcomes but only one, the

first one to happen, is observed; for example, in a study where the interest is to evaluate the death from breast cancer but there are numerous cases of death from other causes, these two outcomes will be “competing” to each other until one of them occur (see more about competing risks in Carvalho et al. (2011) and Klein and Moeschberger (2003)).

For this model, the improper survival function is given by:

$$\begin{aligned}
 S_{pop}(t) &= P(\text{no competent cell is active until time } t) \\
 &= P(N = 0) + P([Z_1 > t, Z_2 > t, \dots, Z_N > t] \cap [N \geq 1]) \\
 &= P(N = 0) + P(Z_1 > t, Z_2 > t, \dots, Z_N > t | N \geq 1)P(N \geq 1) \\
 &= P(N = 0) + \sum_{n=1}^{\infty} P(N = n)(S(t))^n,
 \end{aligned} \tag{2.13}$$

where $S(t)$ is the proper survival function.

This improper survival function is the probability that no competent cell is active until time t because, if this happens it means that the individual has not relapsed or, in other words, the individual survived the relapse. This probability can be divided into two groups, one representing the cured ($N = 0$) individuals and the other, the non-cured ones ($N \geq 1$). In the first case, there is no competent cell that may become active, thus it is given by the probability that $N = 0$; on the other case it will be given by the probability of each of the N , $N \geq 1$, cells has not become active until time t .

As is it usually done (Chen et al., 1999; Sinha et al., 2003; Lambert and Thompson, 2007), in this work the latent random variable N will follow a *Poisson* distribution with mean θ . By doing so, it is possible to obtain that:

$$\begin{aligned}
 S_{pop}(t) &= \exp\{-\theta\} + \sum_{k=1}^{\infty} S(t)^k \frac{\theta^k \exp\{-\theta\}}{k!} \\
 &= \exp\{-\theta F(t)\},
 \end{aligned} \tag{2.14}$$

where $S(t)$ is the genuine survival function. Also, note that $S_{pop}(\infty) = \exp\{-\theta\} \in (0, 1)$.

Other distributions can be assumed for the number of competing cells. de Castro et al. (2009), for example, assumes that N follows a *Negative Binomial* distribution, while Barreto-Souza (2015) considers that this random variable belongs to a mixed *Poisson* class of distributions, which includes the *Negative Binomial* and the *Poisson – Inverse Gaussian* distributions as a particular cases and also takes into account overdispersion.

The improper p.d.f. and hazard function can be easily obtained by using the relationships between these and the survival function:

$$f_{pop}(t) = \theta f(t) \exp\{-\theta F(t)\}, \quad (2.15)$$

$$h_{pop}(t) = \theta f(t). \quad (2.16)$$

Other interesting and useful quantities are the survival, hazard and probability density functions for the non-cured individuals. They are given by:

$$S_{NC}(t) = P(T > t | N \geq 1) = \frac{\exp\{-\theta F(t)\} - \exp\{-\theta\}}{1 - \exp\{-\theta\}}, \quad (2.17)$$

$$f_{NC}(t) = \left(\frac{\exp\{-\theta F(t)\}}{1 - \exp\{-\theta\}} \right) \theta f(t), \quad (2.18)$$

$$h_{NC}(t) = \left(\frac{\exp\{-\theta F(t)\}}{\exp\{-\theta F(t)\} - \exp\{-\theta\}} \right) h_{pop}(t). \quad (2.19)$$

Note that $S_{NC}(t)$ is a proper survival function because $S_{NC}(0) = 1$ and $S_{NC}(\infty) = 0$; consequently $f_{NC}(t)$ and $h_{NC}(t)$ are also proper. From these functions it is possible to obtain an expression that links the mixture and the promotion time models. Thus, from one model it is possible to reach the other (more details in Ibrahim et al. (2001b)).

The likelihood function based on a sample of n individuals with all the information, that is, the observable and non-observable data, is given by:

$$\begin{aligned} L(\Phi, \psi; D) &= \left(\prod_{i=1}^n (S(t_i | \Phi))^{N_i - \delta_i} (N_i f(t_i | \Phi))^{\delta_i} \right) \prod_{i=1}^n \frac{\theta_i^{N_i} \exp\{-\theta_i\}}{N_i!} \\ &= \left(\prod_{i=1}^n (S(t_i | \Phi))^{N_i} (N_i h(t_i | \Phi))^{\delta_i} \right) \prod_{i=1}^n \frac{\theta_i^{N_i} \exp\{-\theta_i\}}{N_i!}, \end{aligned} \quad (2.20)$$

where Φ is the vector of the parameters that indexes the p.d.f., which may include a vector of coefficients. Furthermore, analogously to the mixture model, the probability of cure can be modeled through covariates. In this case, a link function will be used, for example: $\theta_i = \exp\{\mathbf{z}_i \boldsymbol{\psi}\}$. Again, the covariates used to explain the cure fraction can be different than the ones used to model the non-cure individuals, therefore, the entire data is composed by $D = (t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i, N_i)$, $i = 1, 2, \dots, n$, where \mathbf{x}_i represents the covariates vector used to model the i -th non-cured individual and \mathbf{z}_i represents cure probability for the same i -th individual.

As stated before, the probability of cure is given by the probability that the individual has no competent cell: $P(N = 0) = \exp\{-\theta\} = \exp\{-\exp\{\mathbf{z}\boldsymbol{\psi}\}\} \equiv \lim_{t \rightarrow \infty} S_{pop}(t)$. Note that, as $\theta \rightarrow 0$, $P(N = 0) \rightarrow 1$ and as $\theta \rightarrow \infty$, $P(N = 0) \rightarrow 0$.

It is also possible to obtain a closed expression for the likelihood function based only on the observed data, $D_{obs} = (t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$, $i = 1, 2, \dots, n$. In this case, it is necessary to sum out the latent variable \mathbf{N} :

$$\begin{aligned} L(\Phi, \psi; D_{obs}) = \sum_{\mathbf{N}} L(\Phi, \psi; D) &= \prod_{i=1}^n (\theta_i f(t_i | \Phi))^{\delta_i} \exp \{-\theta_i (1 - S(t_i | \Phi))\} \quad (2.21) \\ &= \prod_{i=1}^n (\theta_i h(t_i | \Phi) S(t_i | \Phi))^{\delta_i} \exp \{-\theta_i (1 - S(t_i | \Phi))\}. \end{aligned}$$

Lastly, in counterpart to the mixture model, the prior distribution of the coefficients for the cured fraction may be proper or improper; by using the Promotion Time model with $N_i \sim Poisson(\theta_i)$, for $i = 1, 2, \dots, n$, it is guaranteed that the posterior distribution will be proper (Ibrahim et al., 2001b). Besides that, according to Rodrigues et al. (2009), it is also guaranteed that the condition of proportional hazards is preserved if the distribution of the number of competent cells (N) is the *Poisson* distribution.

3 The Piecewise Exponential Model

In this chapter, the Piecewise Exponential model (PEM) will be introduced. The aim is to discuss about this model's properties and particularities, showing the advantages and the challenges that come along with it, motivating its usage and understanding. It will also be shown how to include covariates and incorporate the cure fraction.

The PEM was proposed by Kalbfleisch and Prentice (1973) and it has been very explored in the literature, mainly focusing on survival analysis.

In order to specify the PEM one has to, at first, consider a partition of the time axis. Thus, to divide the time axis in b intervals, let $\tau = \{s_0, s_1, \dots, s_b\}$ represent the cuts of the intervals, where $0 = s_0 < s_1 < \dots < s_b < \infty$. In that way the intervals will be $I_1 = (s_0, s_1]$, $I_2 = (s_1, s_2]$, \dots , $I_b = (s_{b-1}, s_b]$. After that, a constant failure rate, λ_j , for $j = 1, 2, \dots, b$ is assumed to each interval. So the hazard function is:

$$h(t) = \lambda_j, \text{ for } t \in I_j, j = 1, 2, \dots, b. \quad (3.1)$$

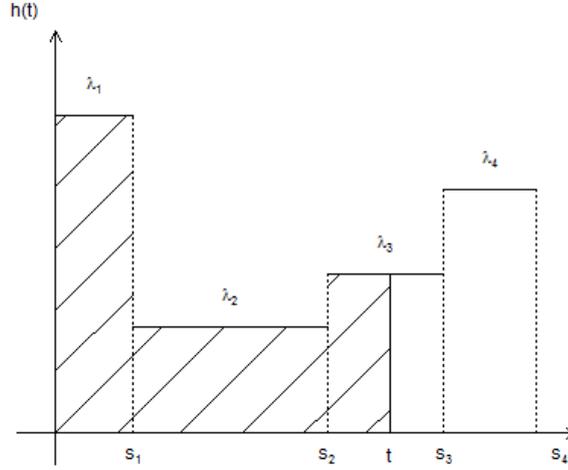
By using such model, one is approximating the hazard function, a continuous function, by line segments; therefore, this function does not have a predetermined shape and, in counterpart of the usual models such as Exponential, Weibull and Log-Normal, no shape must be imposed. This characteristic provides great flexibility for modeling the hazard function and for this reason, the PEM is often considered as a non-parametric model, although in fact, it is a parametric one.

Let T be a non-negative random variable representing the time to event. To introduce the cumulative hazard, consider $t_j, j = 1, 2, \dots, b$ as:

$$t_j = \begin{cases} s_{j-1}, & \text{if } t < s_{j-1} , \\ t, & \text{if } t \in (s_{j-1}, s_j] , \\ s_j, & \text{if } t > s_j . \end{cases} \quad (3.2)$$

For a better comprehension of the usefulness of the quantity t_j , consider a individual whose time to event is represented by Figure 3.1.

Assume that the time grid was divided in four intervals: $I_1 = (0, s_1]$, $I_2 = (s_1, s_2]$, $I_3 = (s_2, s_3]$ and $I_4 = (s_3, s_4]$; suppose also that an individual has suffered the

Figure 3.1: Illustration of the quantity t_j .

event or the censorship at time t , that is, between s_2 and s_3 . The hazard function (λ_1 , λ_2 , λ_3 and λ_4) is the straight line segments and j represents the j -th interval. This time to event is greater than the upper limit of the first interval (s_1), that is $t > s_1$, and therefore $t_1 = s_1$ and this whole interval will be taken in consideration. The same is true for the second interval, therefore $t_2 = s_2$. In turn, the time to event is lower than the upper limit of the third interval (s_3), this means that $t \in (s_2, s_3]$, then $t_3 = t$ so that only a part of it will be considered. Lastly, the time to event is lower than the lower limit of the fourth interval, so $t_4 = s_3$ and the fourth interval will not be taken in consideration.

Given this, the cumulative hazard function can be defined as:

$$H(t|\boldsymbol{\lambda}) = \sum_{j=1}^b \lambda_j(t_j - s_{j-1}), \quad (3.3)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_b)$ is the vector of failure rates.

One may visualize this hazard as the area of each interval. In the case of the Figure 3.1, it follows that

$$\begin{aligned} H(t|\boldsymbol{\lambda}) &= \sum_{j=1}^4 \lambda_j(t_j - s_{j-1}) = \lambda_1(t_1 - s_0) + \lambda_2(t_2 - s_1) + \lambda_3(t_3 - s_2) + \lambda_4(t_4 - s_3) \\ &= \lambda_1(s_1 - s_0) + \lambda_2(s_2 - s_1) + \lambda_3(t - s_2) + \lambda_4(s_3 - s_3) \\ &= \lambda_1(s_1 - s_0) + \lambda_2(s_2 - s_1) + \lambda_3(t - s_2). \end{aligned}$$

Using the relationship between the cumulative hazard function and the survival function, as well as the relationship between the survival function and the probability distribution function, described on page 16, it follows that:

$$S(t|\boldsymbol{\lambda}) = \exp \left\{ - \sum_{j=1}^b \lambda_j (t_j - s_{j-1}) \right\} \quad (3.4)$$

and

$$f(t|\boldsymbol{\lambda}) = \lambda_j \exp \left\{ - \sum_{j=1}^b \lambda_j (t_j - s_{j-1}) \right\}, t \in I_j, \lambda_j > 0, j = 1, 2, \dots, b. \quad (3.5)$$

An important discussion is about the number of intervals. This number can be fixed, as seen in Gamerman (1994), Yin and Ibrahim (2005), de Castro et al. (2009), among others. Nevertheless stipulating the number of intervals is a difficult task. If this number is too large, there will be few data in each interval, therefore it can result in poor and/or unstable estimates; on the other hand, if this number is too small, the hazard function may not be well approximated. Thus, the number of intervals should be carefully chosen, balancing the quantity of data in each interval so that it is possible to provide a good estimation for the hazard function and for the survival function as well. One way to solve this issue would be to estimate the time grid τ , that is, to consider the partition of the time axis and, consequently the number of intervals itself, as an unknown quantity to be estimated. In this work, in the same way as others (Kim et al., 2007; Demarqui, 2010) both cases will be considered. One restriction that may be done is to establish the maximum number of intervals as the number of distinct observed failures so that is guaranteed to exist at least one failure at each interval (Gamerman, 1994).

In order to write the likelihood function one must include the information from all the n individuals, like Equation (2.7), as well as all the b intervals. The likelihood function is given by:

$$\begin{aligned} L(\boldsymbol{\lambda}; D) &= \prod_{j=1}^b \prod_{i=1}^n \lambda_j^{\delta_{ij}} \exp \{ -\lambda_j (t_{ij} - s_{j-1}) \} \\ &= \prod_{j=1}^b \lambda_j^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ -\lambda_j \sum_{i=1}^n (t_{ij} - s_{j-1}) \right\} \\ &= \prod_{j=1}^b \lambda_j^{\eta_j} \exp \{ -\lambda_j \xi_j \}, \end{aligned} \quad (3.6)$$

where $D = \{t_{ij}, \delta_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, b\}$ represents all the data available. In turn t_{ij} is the time of the i -th individual on the j -th interval, δ_{ij} is an indicator variable:

$\delta_{ij} = 1$ if the i -th individual has failed in the j -th interval and $\delta_{ij} = 0$ otherwise. The quantity η_j represents the number of failures, while ξ_j is the total time under test, both associated with the j -th interval.

It is important to notice that estimation by intervals using information only on a specific interval may lead to poor estimates due to few data (Gamerman, 1994) and, in order to solve this issue, the λ 's will be correlated in a way that the information of the actual interval contains the information of the previous one, that is, a dynamic approach will be used.

On the mentioned approach there are some important distributions, namely: the prior, the online and the smoothed distributions. The online distribution, as it is often called, is the posterior distribution for each interval's failure rate, that is, the distribution of the hazard rate of the j -th interval based on all the information until that specific interval, this information will be represented by D_j and the posterior distribution of λ_j will be denoted by $\lambda_j|D_j$. In turn, the smoothed distribution is that one that takes into account all the available information, it will be represented by $\lambda_j|D$. These distributions will be explained ahead.

Note that the the likelihood function (Equation (3.6)), as a function of $\boldsymbol{\lambda}$, corresponds to a product of kernels of *Gamma* distributions with respect to each component of $\boldsymbol{\lambda}$. It means that, if is a *Gamma* prior distribution is considered for the components of the failure rate, conjugacy is obtained. This fact will be essential in this work, both for the parameters' estimation as to the computational aspects, especially when the time grid is estimated. Therefore the *Gamma* distribution will be chosen as the prior distribution of each failure rate λ_j , $j = 1, \dots, b$. Another advantage of eliciting this specific prior is that the *Gamma* distribution is very flexible and it can assume a quite reasonable number of shapes.

Following the dynamic approach proposed by Gamerman (1994), denote the prior information available at the beginning of the study by D_0 . Then, the prior distribution of λ_1 , that is, $\lambda_1|D_0$ is *Gamma*(α_0, γ_0), where α_0 and γ_0 are known values. Uniting the prior information with the likelihood information, it is possible to obtain the posterior distribution of λ_1 , that is, $\lambda_1|D_1$. So, the posterior distribution for the failure rate associated with the first interval is given by:

$$\begin{aligned}
p(\lambda_1|D_1) &\propto L(\boldsymbol{\lambda}; D)p(\lambda_1|D_0) \\
&\propto \lambda_1^{\eta_1+\alpha_0-1} \exp\{-\lambda_1[\gamma_0 + \xi_1]\}.
\end{aligned} \tag{3.7}$$

That is, the posterior distribution of λ_1 is *Gamma*(α_1, γ_1), where $\alpha_1 = \alpha_0 + \eta_1$, $\gamma_1 = \gamma_0 + \xi_1$, and $\eta_1 = \sum_{i=1}^n \delta_{i1}$, $\xi_1 = \sum_{i=1}^n (t_{i1} - s_0)$. As aforementioned, the quantity η_1 represents the total number of failures at the first interval and ξ_1 represents the total time under test at the first interval.

For the second interval there is the initial information, D_0 , and the information from the first interval, this information will be represented by D_1 . Then the prior distribution of failure rate associated to the second interval is $(\lambda_2|D_1, \phi) \sim \text{Gamma}(\phi\alpha_1, \phi\gamma_1)$, where α_1 and γ_1 are the parameters of form and scale of the posterior distribution of λ_1 and ϕ is the discount factor. The discount factor is a number such that, $0 < \phi \leq 1$ and its role is to control the information that is passed successively through the intervals. Consequently, $(\lambda_j|D_{j-1}, \phi) \sim \text{Gamma}(\phi\alpha_{j-1}, \phi\gamma_{j-1})$ for $j = 2, \dots, b$.

Similarly, the posterior distribution for λ_j , $j = 2, \dots, b$, is given by:

$$\begin{aligned}
p(\lambda_j|D_j, \phi) &\propto L(\boldsymbol{\lambda}; D)p(\lambda_j|\phi, D_{j-1}) \\
&\propto \lambda_j^{\eta_j+\phi\alpha_{j-1}-1} \exp\{-\lambda_j[\phi\gamma_{j-1} + \xi_j]\}.
\end{aligned} \tag{3.8}$$

Therefore, $(\lambda_j|D_j, \phi) \sim \text{Gamma}(\alpha_j, \gamma_j)$, where $\alpha_j = \eta_j + \phi\alpha_{j-1}$, $\gamma_j = \phi\gamma_{j-1} + \xi_j$, and $\eta_j = \sum_{i=1}^n \delta_{ij}$, $\xi_j = \sum_{i=1}^n (t_{ij} - s_{j-1})$ for $j = 2, \dots, b$. Likewise the first interval, the quantity η_j represents the total number of failures at the j -th interval and ξ_j represents the total time under test at the j -th interval.

Note that $(\lambda_1|D_0) \equiv (\lambda_1|D_0, \phi)$, it means that the distribution of the initial state (the first interval) does not depend on ϕ , once this quantity begin to be necessary to control the passage from the first interval to the second interval.

The discount factor can be stipulated like Gamerman (1994) and Demarqui (2010) or it can be estimated. In this work it will be estimated. The disadvantage of stipulating ϕ is the requirement of a sensibility study to guarantee a good choice for it; while the estimation of this quantity provides point and interval estimates. It is noteworthy that when ϕ is close to 1, more information is passed through the successive

intervals; if ϕ is equal to 1, all information is passed. On the other hand, when ϕ is close to 0, less information is passed. Moreover, by using a discount factor in this way, the expectation of λ_j is maintained and its variance is inflated (Gamerman, 1994). This characteristic is demonstrated for a general interval j by the following equations:

$$E(\lambda_j|D_{j-1}, \phi) = \frac{\phi\alpha_{j-1}}{\phi\gamma_{j-1}} = \frac{\alpha_{j-1}}{\gamma_{j-1}} = E(\lambda_{j-1}|D_{j-1}) \quad (3.9)$$

and

$$Var(\lambda_j|D_{j-1}, \phi) = \frac{\phi\alpha_{j-1}}{(\phi\gamma_{j-1})^2} = \frac{\alpha_{j-1}}{\phi(\gamma_{j-1})^2} = \frac{1}{\phi}Var(\lambda_{j-1}|D_{j-1}). \quad (3.10)$$

The following step by step will give a better explanation of dynamic scheme proposed by Gamerman (1994) of the prior and posterior distribution of λ :

1. Establish the prior distribution for the failure rate associated to the first interval by choosing the values for α_0 and γ_0 , that is, fully specify $\lambda_1|D_0 \sim Gamma(\alpha_0, \gamma_0)$;
2. Update the prior information with the information from the likelihood, that is, obtain the posterior distribution for the failure rate associated to the first interval, $\lambda_1|D_1 \sim Gamma(\alpha_1, \gamma_1)$;
3. The prior distribution for λ_2 , that is, the failure rate associated to the second interval will be the posterior distribution for λ_1 weighted by the discount factor: $(\lambda_2|D_1, \phi) \sim Gamma(\phi\alpha_1, \phi\gamma_1)$;
4. Obtain the posterior distribution of λ_2 ;
5. For the $j - th$ interval, the prior distribution for λ_j will be the posterior distribution of λ_{j-1} weighted by the discount factor: $(\lambda_j|D_{j-1}, \phi) \sim Gamma(\phi\alpha_{j-1}, \phi\gamma_{j-1})$;
6. Obtain the posterior distribution for the failure rate associated to the $j - th$ interval;
7. Steps 5 and 6 will be repeated until the posterior distribution of failure rate associated to the the last interval $(\lambda_b|D_b)$ is obtained.

An important point of the dynamic approach is the smoothing process, that is, the distribution of λ based on all available information. Gamerman (1994) correlates the components of λ in the log scale, in the following way: $\log(\lambda_j) = \log(\lambda_{j-1}) + w_j$, where w_j is some random disturbance with zero mean and variance W_j . This author

justifies this form of evolution by arguing that it reduces skewness and avoids negative values for λ . By using this specific form of correlation the author obtains an approximated smoothed distribution by making use of linear Bayesian methods (see more in West and Harrison (1997)). This methodology was developed in the context of standard dynamic linear models and requires the first and second order moments, however it only provides estimates for these moments.

Gamerman (1991) and Demarqui (2010) also obtain an approximated smoothed distribution. Other works, such as Kim et al. (2007) and de Castro et al. (2009) also correlated the components of λ in the log scale, but using a different structure to correlate the vector λ .

Another way to correlate the components of λ is by following the approach proposed by Gamerman et al. (2013). This approach was originally developed in time series context although it can be applied to every context that fits the four assumptions described in their article, which includes the PEM as highlighted by the authors themselves. One of the assumptions of the mentioned work is that the evolution equation is of the form: $\lambda_{j+1} = \phi^{-1}\lambda_j\varsigma_{j+1}$, where $\varsigma_{j+1}|D_j, \phi \sim \text{Beta}(\phi\alpha_j, (1-\phi)\alpha_j)$, D_j is the information until the j -th interval and ϕ is the discount factor. For comparison purposes, one may look at this evolution as $\log(\lambda_{j+1}) = \log(\lambda_j) + \varsigma_{j+1}^*$, where $\varsigma_{j+1}^* = \log\left(\frac{\varsigma_{j+1}}{\phi}\right)$.

This last evolution equation introduced will be the one used in the present work. The great advantage in doing so is that, differently from what is mostly found in the literature, exact quantities can be obtained, like the smoothed distribution that will be explained ahead.

The smoothing process is based on the following proposition:

Proposition 3.0.1. *The joint distribution of $(\lambda|\phi, D_b)$ has density given by*

$$p(\lambda|\phi, D_b) = p(\lambda_b|\phi, D_b) \prod_{j=1}^{b-1} p(\lambda_j|\lambda_{j+1}, \phi, D_b)p(\phi|D_b),$$

where the distribution of $(\lambda_j|\lambda_{j+1}, \phi, D_b)$ can be obtained via

$$\lambda_j - \phi\lambda_{j+1}|\lambda_{j+1}, \phi, D_b \sim \text{Gamma}((1-\phi)\alpha_j, \gamma_j), j = 1, 2, \dots, b. \quad (3.11)$$

So, to obtain a value of the smoothed distribution based on Proposition 3.0.1, one has to follow the algorithm given below:

1. Set $j = b$ and sample $p(\lambda_b|\phi, D_b)$ using Proposition 3.0.1;
2. Set $j = j - 1$ and sample $p(\lambda_j|\lambda_{j+1}, \phi, D_b)$ using Equation (3.11);
3. If $j > 1$, go back to step 2; otherwise, the sample of $(\lambda_1, \dots, \lambda_b|\phi, D_b)$ is complete.

The idea then, is to obtain a value of λ_j in a non direct way, through the distribution of $(\lambda_j - \phi\lambda_{j+1}|\lambda_{j+1}, \phi, D_b)$, for $j = 1, \dots, b - 1$, once the distribution of $\lambda_b|D_b$ is already known.

At first, consider a scenario with no covariates and a fixed time grid. In this case, one may wish to estimate $\boldsymbol{\lambda}$ and ϕ . A Bayesian analysis will be performed, so consider the prior distributions that were already described and $\phi \sim \text{Beta}(\theta_1, \theta_2)$. Therefore the joint posterior distribution of $(\boldsymbol{\lambda}, \phi)$ is

$$p(\boldsymbol{\lambda}, \phi|D) \propto L(\boldsymbol{\lambda}; D)p(\boldsymbol{\lambda}|\phi)p(\phi). \quad (3.12)$$

The likelihood function is defined by the Equation (3.6) and the prior distributions were described above. Therefore, $p(\boldsymbol{\lambda}, \phi|D)$ is fully specified. Its calculation is given in the Appendix.

If one wishes to obtain the marginal posterior distribution of ϕ , that is, $\phi|D$, one has to, simply, integrate the joint posterior distribution (Equation (3.12)) with respect to $\boldsymbol{\lambda}$. This calculation as well as the calculation of the full conditional distribution of $\boldsymbol{\lambda}$ are also in the Appendix.

3.1 Random Time Grid

It was previously discussed that the time grid of the PEM can be fixed or estimated. Furthermore, the time grid and the number of intervals have an important role in the modeling procedure. Ibrahim et al. (2001b), for example, states that in the cure rate models context, the number of intervals affects the estimation of the cure rate. This statement accentuates even further the importance of a good choice for the grid.

The fixed time grid may be based on previous experience or it can also be chosen as the number of the distinct failure times Gamerman (1994), for example. Nevertheless, the number of distinct failures may be too large for the number intervals, and

moreover, some of these intervals may be similar to one another, so it would be reasonable to think about grouping them. It can also be seen as allowing the data to indicate which grid is the best one to fit the data.

In this present work the time grid will be estimated via the clustering structure of the Product Partition Model (PPM). The PPM was proposed by Hartigan (1990) and Barry and Hartigan (1992) extended to the case of change point problems. By using this approach it is possible to obtain important information such as the most likely grid *a posteriori*, for example.

The basic idea is to establish a grid with the maximum number of intervals admitted *a priori*, say m' , and, from that, to evaluate the possibility of grouping them. The original grid in this master thesis, also called the finest grid, will have its endpoints as different failure times and it will be set in the following way: after choosing the maximum number of intervals, the number of distinct failures will be equally divided into the m' intervals and the remain ones, if there are some, will be set at the last intervals (from back to front). The purpose of proceeding in such manner is that, generally, at the end of the study there are fewer individuals at risk, then this would be a practical way of allowing more information in the last intervals. For example, in a data set in which there are 23 distinct failures and the maximum number of intervals considered *a priori* is $m' = 9$; in this case there will be 2 distinct failures for the first four intervals and 3 distinct failures for the five remaining ones.

In order to estimate the time grid, consider $\tau' = \{0, y'_1, \dots, y'_m\}$ as the finest grid admitted *a priori*, where y'_1, \dots, y'_m are distinct observed failure times representing the endpoints of the intervals. The vector τ' induces a set of intervals, in this case, $I_1 = (0, y'_1]$, $I_2 = (y'_1, y'_2]$, \dots , $I_{m'} = (y'_{m-1}, y'_m]$. Also, denote by $\mathbb{I} = \{1, \dots, m'\}$ the set of indexes associated to the initial intervals $I_1, \dots, I_{m'}$ and let $\rho = \{i_0, i_1, \dots, i_b\}$, $0 = i_0 < i_1 < \dots < i_b = m'$ be the random partition of \mathbb{I} , which divides the m' initial intervals into b new intervals.

The new intervals will be formed by grouping the original ones. This will be in the form:

$$I_{(\rho)_j} = \cup_{r=i_{j-1}+1}^{i_j} I_r, \quad j = 1, 2, \dots, b. \quad (3.13)$$

The index (ρ) indicates the set of intervals induced by the new partition. By opening expression (3.13) it is possible to see that:

$$\begin{aligned}
I_{(\rho)_1} &= \cup_{r=i_0+1}^{i_1} I_r = I_{i_0+1} \cup \dots \cup I_{i_1} = (0; y'_1]; \\
I_{(\rho)_2} &= \cup_{r=i_1+1}^{i_2} I_r = I_{i_1+1} \cup \dots \cup I_{i_2} = (y'_1; y'_2]; \\
&\vdots \\
I_{(\rho)_b} &= \cup_{r=i_{b-1}+1}^{i_b} I_r = I_{i_{b-1}+1} \cup \dots \cup I_{i_b} = (y'_{i_{b-1}}; y'_{i_b}].
\end{aligned}$$

Note that the elements of the set ρ are the indexes of failure times of the new intervals. Moreover, given the random partition ρ , it is assumed that:

$$h(t) = \lambda_r \equiv \lambda_{(\rho)_j}, \quad (3.14)$$

it means that each of the failure rates associated to the intervals that were united are equal in distribution. The scheme in the Figure 3.2 illustrates the step-by-step that is performed to estimate the time grid, aiming at providing a better understanding.

$$\begin{aligned}
&\tau' = \{0, y'_1, y'_2, y'_3, y'_4, y'_5\} \\
&\quad \downarrow \\
&\underbrace{I_1 = (0, y'_1]}_{\lambda_1}; \underbrace{I_2 = (y'_1, y'_2]}_{\lambda_2}; \underbrace{I_3 = (y'_2, y'_3]}_{\lambda_3}; \underbrace{I_4 = (y'_3, y'_4]}_{\lambda_4}; \underbrace{I_5 = (y'_4, y'_5]}_{\lambda_5} \\
&\quad \downarrow \\
&\mathbb{I} = \{0, 1, 2, 3, 4, 5\} \rightarrow \rho = \{i_0 = 0, i_1 = 1, i_2 = 3, i_3 = 5\} \\
&\quad \downarrow \\
&\underbrace{I_{(\rho)_1} = (0, y'_1]}_{\lambda_{(\rho)_1}}; \underbrace{I_{(\rho)_2} = (y'_1, y'_3]}_{\lambda_{(\rho)_2}}; \underbrace{I_{(\rho)_3} = (y'_3, y'_5]}_{\lambda_{(\rho)_3}}
\end{aligned}$$

Figure 3.2: Illustration of the intervals' grouping scheme.

In the case of the mentioned figure there were initially five intervals: I_1, I_2, I_3, I_4 and I_5 , with $\{y'_1, y'_2, y'_3, y'_4, y'_5\}$ as endpoints. Therefore, $\mathbb{I} = \{0, 1, 2, 3, 4, 5\}$ and, initially, $h(t) = \lambda_j, t \in I_j$. After that, a random partition of \mathbb{I} was chosen: $\rho = \{i_0, i_1, i_2, i_3\} = \{0, 1, 3, 5\}$. This random partition induces the following new set of intervals:

$$\begin{aligned}
I_{(\rho)_1} &= \cup_{r=i_0+1}^{i_1} I_r = \cup_{r=1}^1 I_r = I_1 = (0, y'_1]; \\
I_{(\rho)_2} &= \cup_{r=i_1+1}^{i_2} I_r = \cup_{r=2}^3 I_r = I_2 \cup I_3 = (y'_1, y'_3]; \\
I_{(\rho)_3} &= \cup_{r=i_2+1}^{i_3} I_r = \cup_{r=4}^5 I_r = I_4 \cup I_5 = (y'_3, y'_5].
\end{aligned} \quad (3.15)$$

Thus, given the random partition, $h(t) = \lambda_{(\rho)_j}$, $t \in I_{(\rho)_j}$ and, in this case, λ_2 and λ_3 are considered to be equal in distribution as well as λ_4 and λ_5 .

According to Barry and Hartigan (1992), the PPM is based on the following assumptions:

- i) The prior distribution of ρ has the product form:

$$p(\rho = \{i_0, i_1, \dots, i_b\}) = K^{-1}c(I_{(\rho)_1})c(I_{(\rho)_2}) \dots c(I_{(\rho)_b}), \quad (3.16)$$

where $K^{-1} = \sum_{\mathcal{C}} c(I_{(\rho)_1})c(I_{(\rho)_2}) \dots c(I_{(\rho)_b})$ is the normalizing constant and \mathcal{C} is the set of all possible partitions of the time grid into b intervals. In turn, $c(\cdot)$ is the prior cohesion, which is a quantity representing the degree of similarity between the intervals that are being grouped;

- ii) Conditional on the partition ρ , the model is conditionally independent and thus, have the product form:

$$p(\boldsymbol{\lambda}_{(\rho)}|\rho = \{i_0, i_1, \dots, i_b\}) = \prod_{j=1}^b p(\boldsymbol{\lambda}_{(\rho)_j}|\rho). \quad (3.17)$$

If there is no information available about the similarity among the intervals, one can simply assume that *a priori* the cohesion is equal to 1 for every single one of them. That is, to use the discrete *Uniform* prior.

In the case of the simplest model, with no covariates and no cure fraction, and considering the prior distribution for ρ mentioned above, the posterior distribution of $(\rho|\phi, D)$ is given by:

$$\begin{aligned} p(\rho|\phi, D) &= \int_{\boldsymbol{\lambda}_{(\rho)}} p(\boldsymbol{\lambda}_{(\rho)}, \rho|\phi, D)d\boldsymbol{\lambda}_{(\rho)} \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho)p(\phi)p(\rho)d\boldsymbol{\lambda}_{(\rho)} \\ &\propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho)d\boldsymbol{\lambda}_{(\rho)} \\ &\propto \frac{(\gamma_0)^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_0 + \eta_1)}{(\gamma_0 + \xi_1)^{\alpha_0 + \eta_1}} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\phi\alpha_{j-1} + \eta_j)}{(\phi\gamma_{j-1} + \xi_j)^{(\phi\alpha_{j-1} + \eta_j)}} \\ &\propto \frac{(\gamma_0)^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_1)}{(\gamma_1)^{\alpha_1}} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\alpha_j)}{(\gamma_j)^{\alpha_j}}. \end{aligned} \quad (3.18)$$

Note that the expression (3.18) can only be obtained due to the conjugacy of $\boldsymbol{\lambda}$, otherwise it would be necessary to use numerical methods to calculate the integral.

The algorithm to estimate the time grid that will be used in this present work is the one proposed by Loschi and Cruz (2005). Following their method, consider the random variable U representing the similarity between the intervals:

$$U_j = \begin{cases} 1, & \text{if } \lambda_j = \lambda_{j+1}; \\ 0, & \text{if } \lambda_j \neq \lambda_{j+1} \end{cases} \quad (3.19)$$

for $j = 1, 2, \dots, b-1$. That is, $U_j = 1$ if the j -th and the $(j+1)$ -th intervals are similar and therefore it is fair to group them; and $U_j = 0$ otherwise.

The idea then, is to compare the intervals, two by two, through the predictive distribution, to verify the similarity between them. If they are similar, they are united into one interval; if they are different, they remain the same (they remain being two different intervals). To generate a sample of U , consider the quantity R_j , for $j = 1, 2, \dots, b-1$:

$$\begin{aligned} R_j &= \frac{p(U_j = 1 | U_1 = u_1, \dots, U_{j-1} = u_{j-1}, U_{j+1} = u_{j+1}, \dots, U_{b-1} = u_{b-1}, D)}{p(U_j = 0 | U_1 = u_1, \dots, U_{j-1} = u_{j-1}, U_{j+1} = u_{j+1}, \dots, U_{b-1} = u_{b-1}, D)} \\ &= \frac{p(D | \rho_1)}{p(D | \rho_0)}, \end{aligned} \quad (3.20)$$

where ρ_0 and ρ_1 represent different partitions, according to $U_j = 0$ and $U_j = 1$, respectively. D represents the data available.

The proposed combination of the intervals will be accepted or not according to the following condition:

$$U_j = \begin{cases} 1, & \text{if } R_j \geq \frac{1-u}{u}; \\ 0, & \text{otherwise.} \end{cases}$$

where $j = 1, 2, \dots, b-1$ and u is a value from the *Uniform*(0, 1) distribution. In the case represented by Figure 3.2, $\mathbf{U} = (U_1, U_2, U_3, U_4) = (0, 1, 0, 1)$.

Once the procedure of estimating the time grid is settled, it is possible to go forward with the estimation of the remaining parameters.

It is important to highlight that the posterior distribution of λ_r , $r = 1, 2, \dots, m'$ will be the following mixture:

$$p(\lambda_r | D) = \sum_{i_{j-1} < r \leq i_j} p(\lambda_{(\rho)_j} | D, \rho) R(I_{(\rho)_j} | D), \quad (3.21)$$

where, $R(I_{(\rho)_j} | D)$ is the posterior relevance, that is, the probability that the j -th new interval appears in the partition induced by ρ . In turn, $p(\lambda_{(\rho)_j} | D, \rho)$ is the full conditional distribution of the failure rates associated to the intervals to be grouped.

Including Covariates into the Model

In a study where there is not only the response variable but also some explanatory variables, it is interesting to include this information into the model for its importance to describe the response.

Assume that there are k explanatory variables (or covariates) available. Such covariates are included into the model in a multiplicative way through the hazard function, that is, $h(t) = \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \}$ for $t \in I_{(\rho)_j}$, where $\boldsymbol{\beta}$ is the coefficients vector and \mathbf{x}_i is the vector of covariates of the i -th individual. Note that the baseline hazard is $h_0(t) = \lambda_{(\rho)_j}$, for $t \in I_{(\rho)_j}$.

In this case, the likelihood is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \rho; D) &= \prod_{j=1}^b \prod_{i=1}^n (\lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \})^{\delta_{ij}} \exp \{ -\lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \} \\ &= \exp \left\{ \sum_{i=1}^n \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ -\lambda_{(\rho)_j} \sum_{i=1}^n [\exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1})] \right\}. \end{aligned} \quad (3.22)$$

The aim here is to estimate, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}_{(\rho)}$, ϕ and ρ . It will be assumed that *a priori* the vector of coefficients $\boldsymbol{\beta}$ does not dependent neither on $\boldsymbol{\lambda}_{(\rho)}$ nor on ϕ or ρ . Based on this, it is possible to calculate some important distributions, such as the joint posterior distribution, the full conditional distributions and others.

The posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho)$ is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho | D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \rho; D) p(\boldsymbol{\beta}) p(\boldsymbol{\lambda}_{(\rho)} | \boldsymbol{\beta}, \phi, \rho) p(\phi) p(\rho). \quad (3.23)$$

The likelihood function as well as the prior distributions of $(\boldsymbol{\lambda}_{(\rho)} | \boldsymbol{\beta}, \phi, \rho)$ and of ϕ were already specified. The prior distribution for ρ , in turn, will be the Bayes-Laplace prior and the vector $\boldsymbol{\beta}$ will be considered independent *a priori*, with a $Normal(0, \sigma_l^2)$ distribution, for $l = 1, \dots, k$.

The parameters of interest, that is, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}_{(\rho)}$, ϕ and ρ will be estimated according to the following expressions:

$$p(\boldsymbol{\beta}|\boldsymbol{\lambda}_{(\rho)}, \phi, \rho, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \rho; D)p(\boldsymbol{\beta}), \quad (3.24)$$

$$p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho), \quad (3.25)$$

$$p(\phi|\boldsymbol{\beta}, \rho, D) \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho)p(\phi)d\boldsymbol{\lambda}_{(\rho)}, \quad (3.26)$$

$$p(\rho|\boldsymbol{\beta}, \phi, D) \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho)p(\rho)d\boldsymbol{\lambda}_{(\rho)}. \quad (3.27)$$

The expressions associated with these distributions are presented in the Appendix. It is noteworthy that the conjugacy for $\lambda_{(\rho)_j}$ is maintained, that is, $\lambda_{(\rho)_j} | (\boldsymbol{\beta}, \phi, \rho, D)$ still follows a *Gamma* distribution (see Equation (A.11)). The full conditional distribution of $\lambda_{(\rho)_j}$ is *Gamma* $(\alpha_0 + \eta_1, \gamma_0 + \xi_1)$ for $j = 1$ and *Gamma* $(\phi\alpha_{j-1} + \eta_j, \phi\gamma_{j-1} + \xi_j)$ for $j = 2, \dots, b$. But at this point, differently from the case with no covariates, $\xi_j = \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{ij} - s_{j-1})$, for $j = 1, 2, \dots, b$ and this quantity no longer represents the total time under test. Nevertheless, even when covariates are included into the model it is still possible to calculate the integral involving the distribution of $(\rho|\boldsymbol{\beta}, \phi, D)$.

Incorporating the Cure Fraction

Studies in which it is plausible to assume that a proportion of subjects will never experience the event of interest, are those which cure fraction models can be applied to. A way of verifying the coherence or necessity of using such method is by verifying if there is a plateau on the Kaplan-Meier estimator, or, in other words, if the estimate becomes constant, in a value greater than zero, at a certain point of the time and remains in that way until the end of the follow-up.

The cure fraction will be incorporated into the model by using the Promotion Time Model and by considering the PEM to describe the promotion times. In this case, substituting Equations 3.4 and 3.5 in Equation 2.20, the likelihood function based on all the information is given by:

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) &= \left(\prod_{j=1}^b \prod_{i=1}^n \exp \{ \lambda_{(\rho)_j} N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \} (N_i \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \})^{\delta_{ij}} \right) \\
&\quad \prod_{i=1}^n \frac{\theta_i^{N_i} \exp \{ -\theta_i \}}{N_i!} \\
&= \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ -\lambda_{(\rho)_j} \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \right) \\
&\quad \left(\prod_{i=1}^n \frac{\theta_i^{N_i} \exp \{ -\theta_i \}}{N_i!} N_i^{\sum_{j=1}^b \delta_{ij}} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \right), \tag{3.28}
\end{aligned}$$

where $\theta_i = \exp \{ \mathbf{z}_i \boldsymbol{\psi} \}$.

In turn, the likelihood function based only on the observed information is:

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D_{obs}) &= \sum_N L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) \\
&= \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \right) \prod_{i=1}^n \theta_i^{\sum_{j=1}^b \delta_{ij}} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \\
&\quad \exp \left\{ -\sum_{j=1}^b \delta_{ij} \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \tag{3.29} \\
&\quad \exp \left\{ -\theta_i \left(1 - \exp \left\{ -\sum_{j=1}^b \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \right) \right\}.
\end{aligned}$$

Therefore, the parameters to be estimated are: $\boldsymbol{\beta}$, $\boldsymbol{\lambda}_{(\rho)}$, ϕ , $\boldsymbol{\psi}$ and ρ . Moreover, once N is a latent variable, it is necessary to generate the number of the competent cells for the n individuals. It can be demonstrated (Ibrahim et al., 2001b) that $(N_i | \boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)_j}, \boldsymbol{\psi}, D_{obs}) \sim \text{Poisson}(S(t_i | \boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}) \exp \{ \mathbf{z}_i \boldsymbol{\psi} \}) + \delta_i$, where δ_i is the indicator of censorship, for $i = 1, 2, \dots, n$.

The posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \boldsymbol{\psi}, \rho)$ is given by:

$$p(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \boldsymbol{\psi}, \rho | D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) p(\boldsymbol{\beta}) p(\boldsymbol{\lambda}_{(\rho)} | \boldsymbol{\beta}, \phi, \rho) p(\phi) p(\boldsymbol{\psi}) p(\rho) \tag{3.30}$$

The likelihood function based on all the information will be used to estimate $\boldsymbol{\lambda}_{(\rho)}$, ϕ , $\boldsymbol{\psi}$ and ρ . In turn, for the vector $\boldsymbol{\beta}$ the likelihood function based only on the observed information will be used. The intention of doing in such way is to improve the

convergence of the parameters by eliminating the uncertainty arising from N . So, the parameters of interest will be estimated according to the following:

$$p(\boldsymbol{\beta}|\boldsymbol{\lambda}_{(\rho)}, \phi, \boldsymbol{\psi}, \rho, D_{obs}) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D_{obs})p(\boldsymbol{\beta}), \quad (3.31)$$

$$p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, \rho, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho), \quad (3.32)$$

$$p(\phi|\boldsymbol{\beta}, \boldsymbol{\psi}, \rho, D) \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho)p(\phi)d\boldsymbol{\lambda}_{(\rho)}, \quad (3.33)$$

$$p(\boldsymbol{\psi}|\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D_{obs})p(\boldsymbol{\psi}), \quad (3.34)$$

$$p(\rho|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, D) \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho)p(\rho)d\boldsymbol{\lambda}_{(\rho)}. \quad (3.35)$$

The expressions related to these distributions can be found on the Appendix.

Note once more that, even when considering a cure fraction, the conjugacy is maintained:

$\lambda_{(\rho)_1} \sim \text{Gamma}(\alpha_0 + \eta_1, \gamma_0 + \xi_1)$ and $\lambda_{(\rho)_j} \sim \text{Gamma}(\phi\alpha_{j-1} + \eta_j, \phi\gamma_{j-1} + \xi_j)$ for $j = 2, \dots, b$ but, at this point, $\xi_j = \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\}(t_{ij} - s_{j-1})$. Thus the the integral related to the distribution of $(\rho|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, D)$ still can be solved.

4 Applications

In order to evaluate the progress of this work, some applications were made. In total there are two applications: the first one illustrates the case of the simple model and the second one, the inclusion of the cure fraction. In both applications the estimates were obtained by using the fixed and the random time grid, so it could be possible to compare the pros and cons of each approach. It is noteworthy that regardless of the methodology applied to the time grid, the discount factor was considered unknown and thus, it was estimated.

The computational methods used in this work were the Gibbs Sampler and the Adaptive Rejection Sampler (ARS) (Gilks and Wild, 1992). The ARS method is used to generate values from a distribution when its expression does not have a closed form. To use the ARS, the functions of interest, for example, the kernel of the full conditional distributions, must be log-concave. By definition, if a function $f(x)$ is log-concave, this means that $f'(x)$ decreases monotonically with increasing x , in its domain. If the function is not log-concave, the Adaptive Rejection Metropolis Sampling within Gibbs Sampling (ARMS) (Gilks et al., 1995) can be used. More information about these methods can be found in Gamerman and Lopes (2006).

All analyzes were performed by using the R software, version 3.1.2 (R Core Team, 2014). The package required to use the command “ars” was the “dlm” package (Petris, 2010). An important argument of this command is the domain of the function. It is well-known that the domain of the coefficients (β and ψ), is $(-\infty, \infty)$; however, to facilitate the computational aspects related to generating values of the full conditional distributions of the coefficients, the logit transformation was applied. That is, $\beta_{trans} = \frac{\exp\{\beta\}}{1 + \exp\{\beta\}}$ and $\psi_{trans} = \frac{\exp\{\psi\}}{1 + \exp\{\psi\}}$, in this way, the domain of the transformation is $(0, 1)$ and can be used in the R function.

The comparison of the models was based on the LPML (Logarithm of the Pseudo-Marginal Likelihood) and the WAIC (Watanabe-Akaike or Widely Applicable Inform) (Watanabe, 2010; Vehtari and Gelman, 2014) criteria. More information about

these methods can be found on the Appendix (page 60). However, following these quantities, the higher they are the better is the model.

4.1 Brain Cancer Data

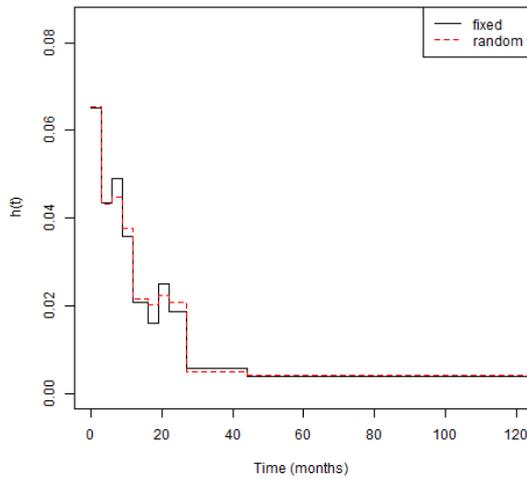
On the first application there was no covariates. In this case, the data utilized proceeds from a brain cancer study obtained from SEER (Surveillance, Epidemiology, and End Results Program) database (<http://seer.cancer.gov/seerstat/>). This data set is composed by the time, in months, to death from brain cancer of 231 individuals. From the total patients, 58.01% deceased.

At first, the failure rate along with the discount factor were estimated with a fixed time grid, and after that, the number of intervals was considered to be unknown and the all the parameters were estimated again.

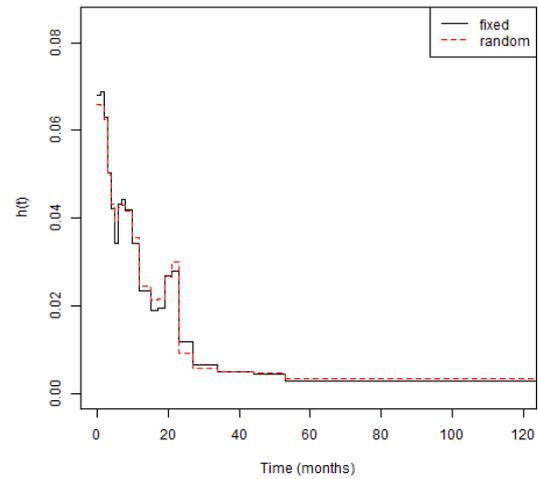
The burn-in considered was 50000, a lag of 100 and 1000 posterior values were obtained, resulting in a total of 150000 iterations. The prior distributions were vague: $(\lambda_{(\rho)_1} | D_0) \sim \text{Gamma}(0.001, 0.001)$ and $\phi \sim \text{Beta}(1, 1)$. In the case of fixed number of intervals, this number was established as 10, 20 and 32, this last value is the number of distinct observed failures. On the other hand, in the case of a random time grid, the same values were considered for the maximum number of intervals. Thus, the estimated number of intervals may vary from 1 to 10, or 1 to 20, or 1 to 32, on the last case.

Figure 4.1 shows the failures rate estimates for the fixed and random time grid, varying the (maximum) number of intervals. The estimates were based on the posterior medians. Note that the estimates based on the random time grid are smoother than those based on the fixed grid.

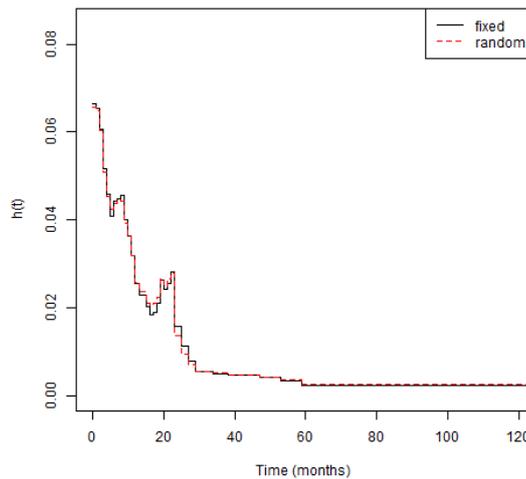
Regarding the discount factor, consider the Figure 4.2. Note that, as the number of intervals increases, ϕ also increases. The explanation for this is: as the number of intervals increases, the less is the information that remains in each interval, it means that the more is the information that will have to be carried along the intervals. Besides that, note that the discount factor for the fixed time grid is higher than that of the random time grid. Such difference occurs because when the time grid is being fixed the discount factor will try to adapt the passage of information for that unique grid; on the other hand, when the time grid is estimated, the data are informing which grid is the best one to fit



(a) 10 intervals.



(b) 20 intervals.



(c) 32 intervals.

Figure 4.1: Comparison of the failure rate estimate varying the number of intervals.

the model, thus if the best grid was obtained it is reasonable to think that the information is set on the most suitable interval, and, therefore, less information will be necessary to be passed through the intervals. Moreover, in this case, the estimated number of intervals will always be lower or equal to the fixed number of intervals, therefore the estimated discount factor in the random time grid scenario will always be lower or equal to the estimated discount factor obtained by fixing the time grid.

Table 4.1 shows the LPML and WAIC results for comparing the models presented; the higher these measures are, the better. The values in bold represent the best

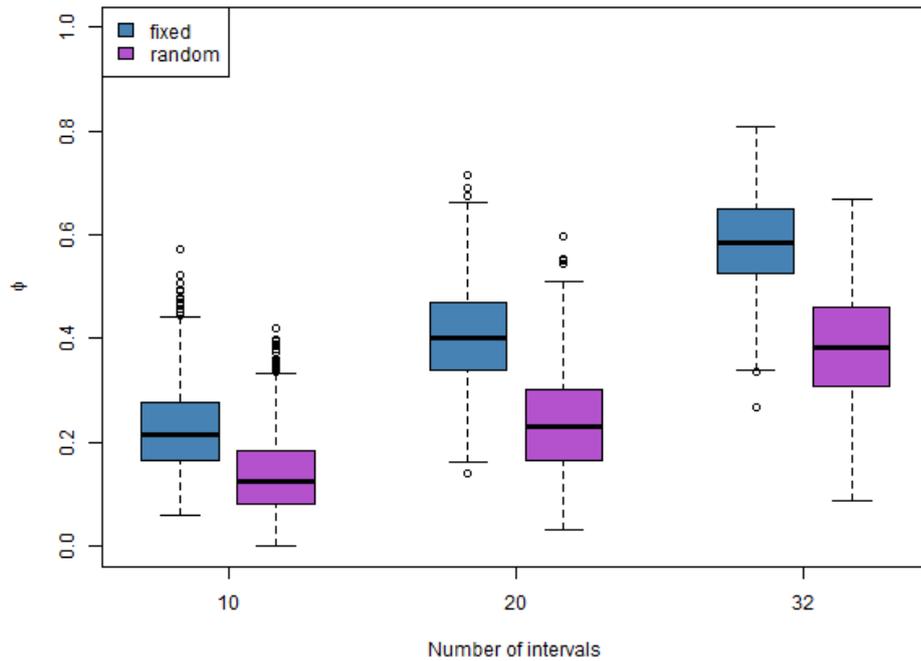


Figure 4.2: Boxplot of the posterior discount factor sample for fixed and random time grid varying the (maximum) number of intervals.

model for each approach (random and fixed grid), according to each comparison criteria. Thus, according to the LPML criteria, the case with fixed time grid with 20 intervals is the best model and, immediately after, follows the case with random time grid with the maximum number of intervals established as 20. On the other hand, by analyzing the WAIC one would conclude the contrary: to estimate the number of intervals (considering $m' = 20$) is better than to fix it at 20 intervals. These results can also be observed through Figure 4.3.

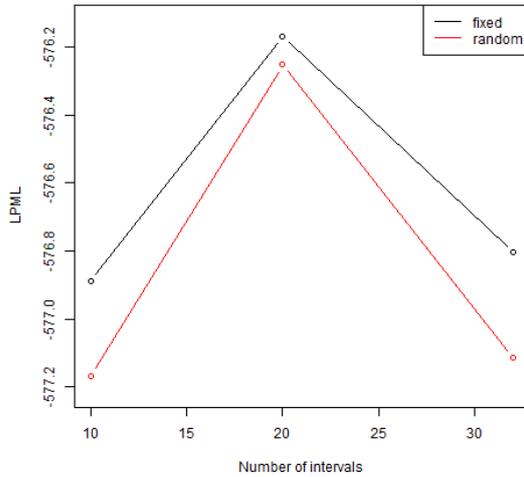
Following both criteria of goodness of fit, there may be a doubt among fixed or random time grid. However, it is worth highlighting important points to consider, such as the information about the number of intervals and the most probable grid. These information can only be obtained by considering the number of intervals as an unknown quantity to be estimated. Therefore the final model will be the one with random time grid and $m' = 20$.

Table 4.2 shows descriptive statistics of the number of intervals for the three cases considered ($m' = 10, 20$ and 32) and Figure 4.4 illustrates the posterior distribution of the number of intervals. Note that, for the case in which $m' = 10$, the maximum value

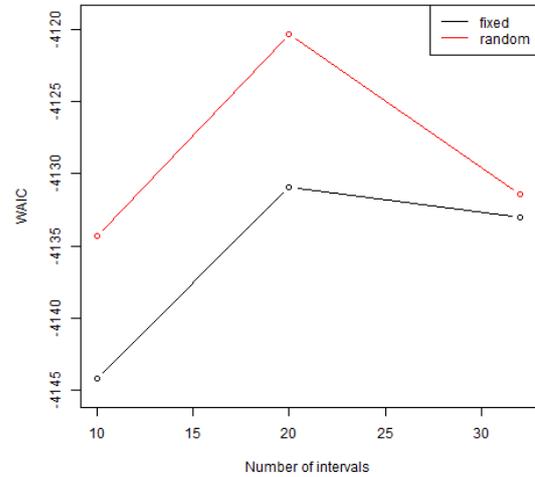
Table 4.1: Comparison of fixed time grid x random time grid according to different (maximum) number of intervals.

Fixed time grid		
Number of intervals	LPML	WAIC
10	-576.8872	-4144.1555
20	-576.1691	-4130.9230
32	-576.8023	-4133.0540

Random time grid		
Maximum number of intervals	LPML	WAIC
10	-577.1669	-4134.3280
20	-576.2522	-4120.3278
32	-577.1141	-4131.4528



(a) LPML.



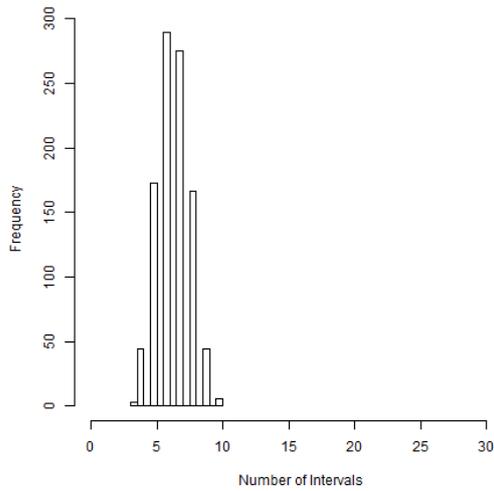
(b) WAIC.

Figure 4.3: Model comparison measures for the fixed time grid x random time grid according to different (maximum) number of intervals.

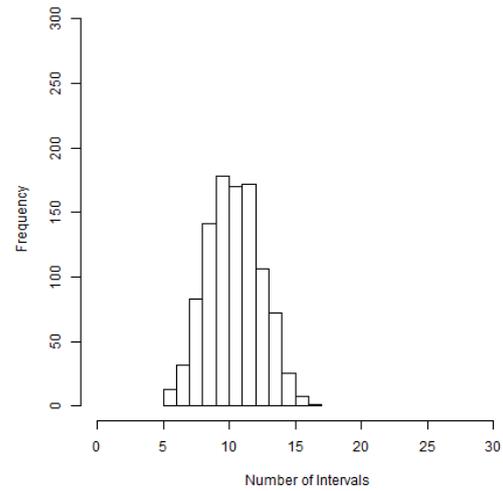
estimated was 10, which may indicate the necessity for more intervals. On the other hand, when $m' = 32$ the mode value was 16 and the third quartile was 19, which may reinforce that the grid with 20 intervals is a good choice.

Table 4.2: Summary of the number of intervals.

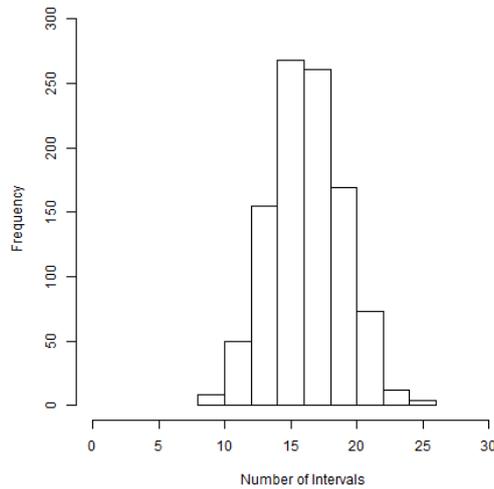
m'	Minimum	1 ^o Quartile	Median	Mean	Mode	3 ^o Quartile	Maximum
10	3	6	6	6.493	6	7	10
20	5	9	11	10.84	10	12	17
32	8	15	17	16.73	16	19	26



(a) 10 intervals.



(b) 20 intervals.



(c) 32 intervals.

Figure 4.4: Histograms of the number of intervals varying the maximum number of intervals.

Based on the results of the chosen model, there were five grids with the higher probability, they are: $\tau_1 = \{0, 3, 5, 6, 10, 12, 17, 21, 23, 27, 53\}$, $\tau_2 = \{0, 4, 10, 12, 15, 21, 23, 27, 44\}$,

$\tau_3 = \{0, 2, 3, 4, 5, 12, 15, 21, 23, 53\}$, $\tau_4 = \{0, 3, 6, 12, 21, 23, 53\}$ and $\tau_5 = \{0, 3, 4, 12, 19, 23, 27, 34\}$, each one of them with probability 0.002 of being sampled. Although this value may seem small, note that the number of possible grids is $2^{m'-1} = 2^{19}$.

From the posterior sample of the chosen model, it is possible to infer that a good number for the intervals is the modal value 10. Besides that, the probability that the number of intervals is between 7 and 14 is 95%, which emphasizes how unnecessary it is to establish this number as 32 intervals, for example. Moreover, the posterior median for the discount factor is 0.2300 and its HPD interval is [0.0716; 0.4343].

In turn, in Figure 4.5 it is possible to compare the estimated survival function. The black line represents the estimate from the Kaplan-Meier estimator and the red line shows the estimated based on the posterior medians obtained by the proposed approach. Note that the estimates are very similar.

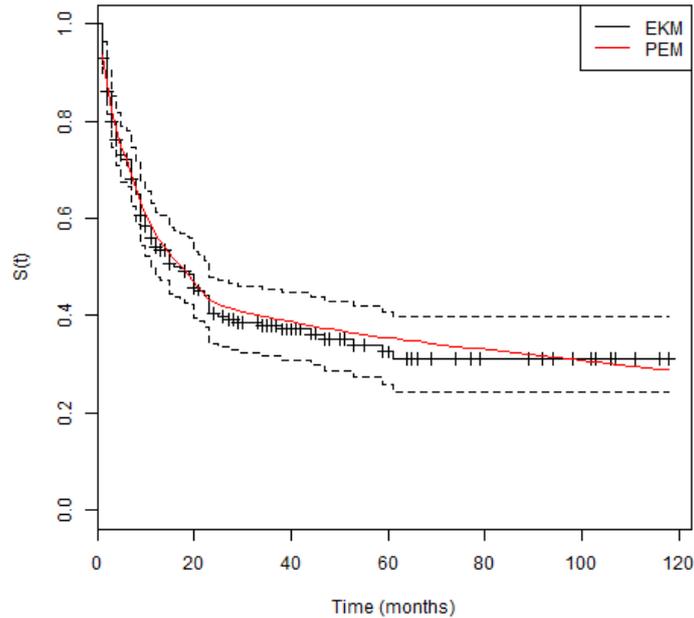


Figure 4.5: Estimated survival function.

Demarqui (2010) obtained considerably similar results when analyzing this data set. In the mentioned work, the best model was also that one with based on the dynamic approach with random time grid and $m' = 20$. The discount factor in Demarqui's work was fixed, which required a sensibility analysis. Within the fixed values, the best model was the one with $\phi = 0.25$ which is really close to the value estimated in the present work (0.23).

4.2 Melanoma Data

The second application is based on the E1673 clinical trial conducted by Eastern Cooperative Oncology Group (ECOG), with 650 patients who had melanoma. These data were taken from Ibrahim et al. (2001b). The response variable is the time until death (in years) from the disease and the covariates are: age (in years), sex (0 - male; 1 - female) and performance status (0 - fully active; 1 - other).

The median age was approximately 49 years; 42.31% of the patients were female and 86.31% patients were fully active. After nearly twenty years of follow-up, 60.46% had suffered the event of interest.

According to Li et al. (2001) the usage of cure fraction models requires a long time of follow-up and also a large number of censorship, which is the case of the E1673 study. In addition, the plateau can be observed in Figure 2.1, note that at the end of follow-up, the estimated survival function remains constant at approximately 0.4.

The aim in this application is to verify which covariates are influencing the cure probability as well as which ones are affecting the non-cured individuals. Moreover, to evaluate the difference between fixed and random time grid, varying the number of intervals and verifying how it affects quantities of interest, such as the hazard function and the cure probability.

In this case there will be two vectors of regression coefficients. The first one will be attached to the failures, in the following way $h(t_i) = h_0(t_i) \exp\{\mathbf{x}_i\boldsymbol{\beta}\} = \lambda_{(\rho)_j} \exp\{x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3\} = \lambda_{(\rho)_j} \exp\{\text{age}_i\beta_1 + \text{sex}_i\beta_2 + \text{PS}_i\beta_3\}$, for $i = 1, 2, \dots, n$ and $t_i \in I_{(\rho)_j}$. The second one, in turn, will be representing the cure fraction, that is, $\theta_i = \exp\{\mathbf{z}_i\boldsymbol{\psi}\} = \exp\{z_{i1}\psi_1 + z_{i2}\psi_2 + z_{i3}\psi_3 + z_{i4}\psi_4\} = \exp\{\psi_1 + \text{age}_i\psi_2 + \text{sex}_i\psi_3 + \text{PS}_i\psi_4\}$, where $i = 1, 2, \dots, n$ and ψ_1 represents the intercept.

Similarly to the first application, the burn-in considered was 50000, a lag of 100 and 900 posterior values were obtained, resulting in 140000 iterations. The prior distributions were also vague: $(\lambda_{(\rho)_1} | D_0) \sim \text{Gamma}(0.001, 0.001)$, $\phi \sim \text{Beta}(1, 1)$, $\beta_l \sim \text{Normal}(0, 1000)$, for $l = 1, 2, 3$ and $\psi_k \sim \text{Normal}(0, 1000)$, for $k = 1, 2, 3, 4$. Besides that, the covariate age was standardized in order to improve convergence procedures.

The possible maximum number of intervals, that is, the number of distinct observed failures was 361 but, as demonstrated by the previous application, this number

may be a too large number to be considered. In this line of thought, the grid was fixed in 10, 20, 30 and 40 intervals. These same values were used for the maximum number of intervals (m').

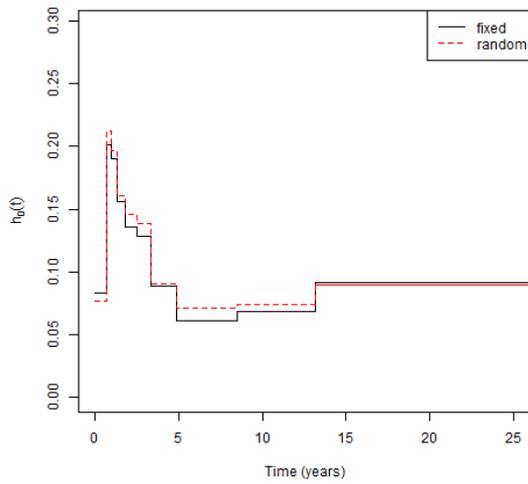
In Figure 4.6 it is possible to observe the estimated baseline hazard function, based on the posterior medians. Note that as the maximum number of intervals increases the estimates become dissonant, comparing to the fixed time grid.

Table 4.3 shows the LPML and WAIC measures for the goodness of fit, the values in bold represent the best model for each approach (random and fixed grid), following the adopted measures of goodness of fit. According to LPML values, the best grid is the grid in the random framework with $m' = 20$ and right after that comes the time grid fixed at 20 intervals. According to the WAIC measure though, the conclusion is different: the best model is that one with random time grid with the maximum number of intervals established at 30; and considering only fixed time grids, it would be best to fix it at 10 intervals. This can also be seen in Figure 4.7.

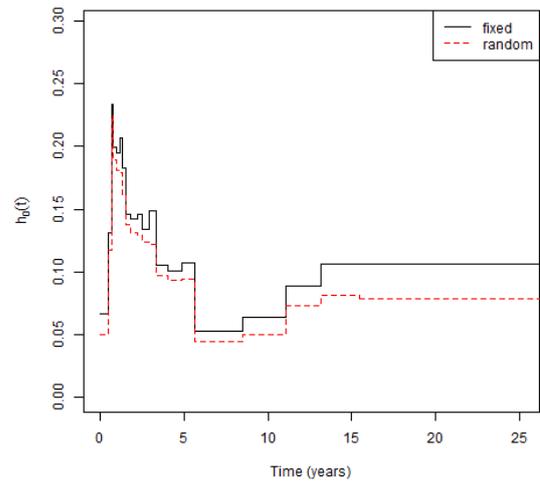
Table 4.3: LPML and WAIC results.

Fixed time grid		
Number of intervals	LPML	WAIC
10	-1341.2429	-2822.4199
20	-1333.8083	-2824.2849
30	-1352.4812	-2869.8333
40	-1391.1048	-2957.5321
Random time grid		
Maximum number of intervals	LPML	WAIC
10	-1341.2722	-2835.1465
20	-1331.0072	-2819.8805
30	-1334.0243	-2813.8030
40	-1338.7437	-2814.0841

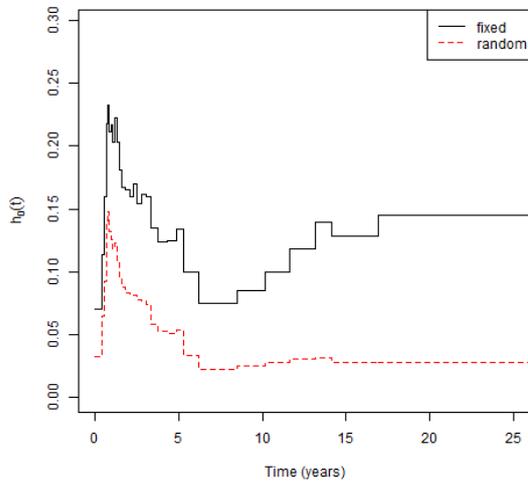
In the works of Demarqui (2010) and Kim et al. (2007) these data were also analyzed using the Promotion Time Model and the PEM for the promotion times. Both works used the LPML criteria and got to the conclusion that it is best to estimate the time



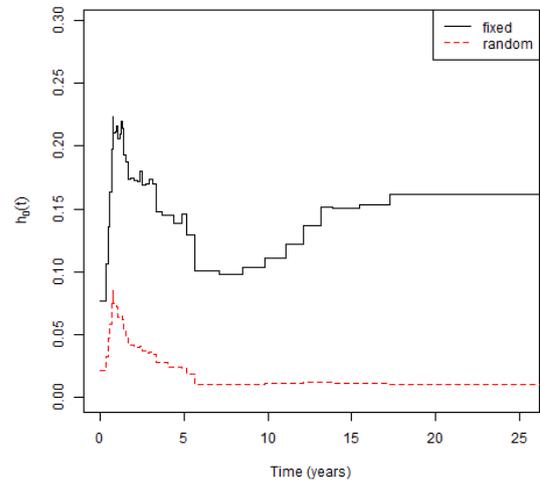
(a) 10 intervals.



(b) 20 intervals.



(c) 30 intervals.



(d) 40 intervals.

Figure 4.6: Comparison of the failure rate estimates using fixed and random time grid and also varying the (maximum) number of intervals.

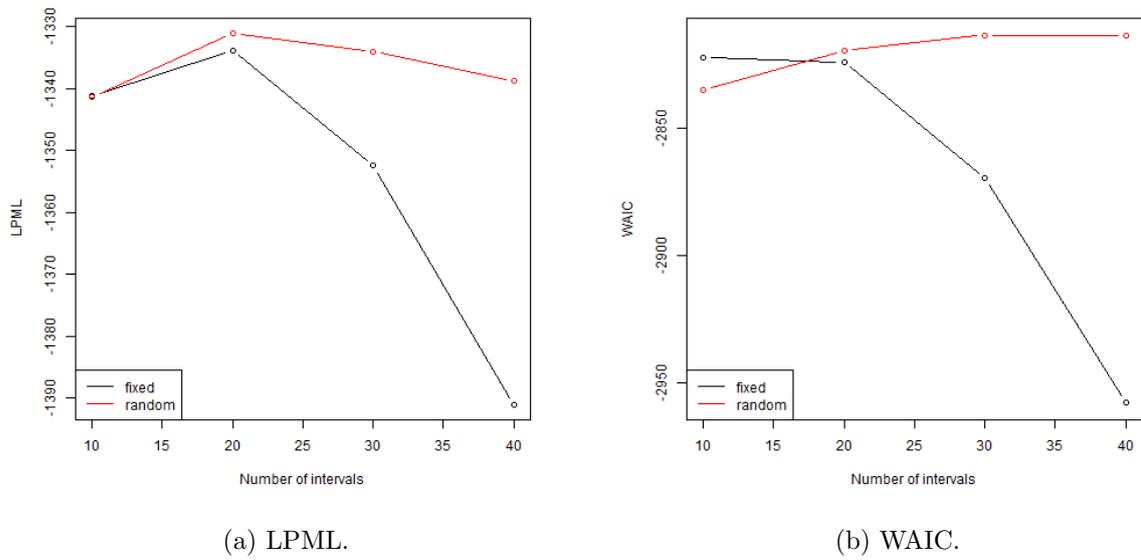


Figure 4.7: Model comparison measures for the fixed time grid x random time grid according to different (maximum) number of intervals.

grid. However, the results obtained by Demarqui (2010) indicates that the best model is that one with $m' = 20$, and the mode value obtained was 10; while the approach used by Kim et al. (2007) (which does not establish a maximum number of intervals) results in a modal value equal to 30.

Nevertheless as it is shown in Table 4.4, it was noted that as the number of maximum number of intervals increases, the estimate of ψ_1 , the intercept to model the cure fraction, become more unstable and dissonant comparing to the others, affecting directly the probability of cure. In Figure 4.8 there is the probability of cure regardless of the significance of the parameters. Note how different are the estimates obtained by the random time grid with $m' = 30$ and $m' = 40$ from the others.

Table 4.4: Estimates of the intercept (ψ_1) for the case with random time grid varying the maximum number of intervals.

m'	Median	Mean	Standard Deviation	HPD 95%
10	0.2846	0.4678	0.5108	[-0.0412; 1.7096]
20	0.3287	0.6592	0.6759	[-0.0098; 2.2392]
30	0.6868	0.9738	0.8005	[0.0552; 2.5060]
40	1.3125	1.2083	0.8788	[0.0858; 2.5614]

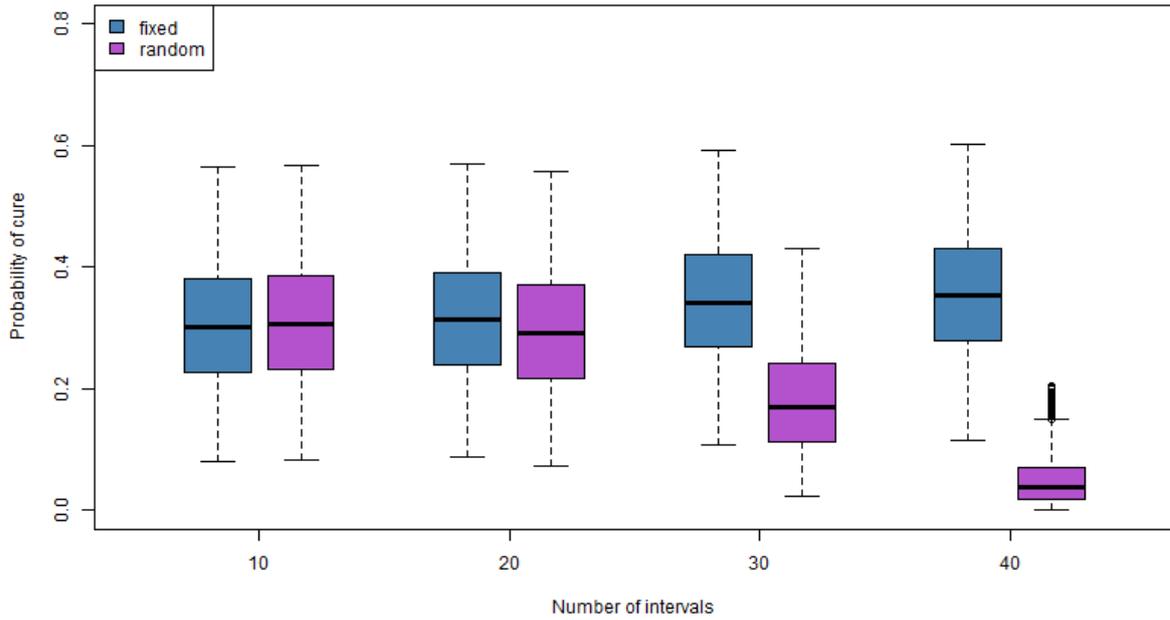


Figure 4.8: Comparison of the probability of cure estimated varying the number of intervals.

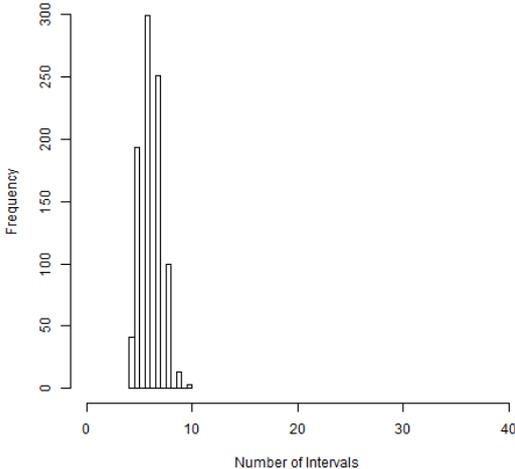
Taking in consideration the points of view expressed above and the similarity between this present work and Demarqui's, the chosen final model will be the one with a random time grid and $m' = 20$.

Table 4.5 shows estimates for the number of intervals, for $m' = 10, 20, 30$ and 40. Figure 4.9 illustrates the histogram of the posterior values for the number of intervals in each specific case. Consider the estimates obtained by choosing $m' = 30$, the third quartile is 18, which is lower than 20, this may be another indicative to choose $m' = 20$ as the final model.

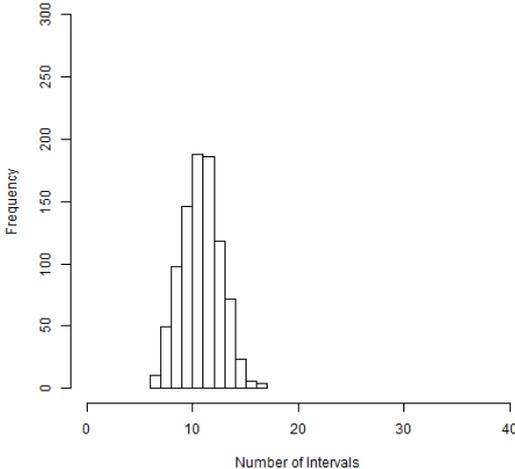
Table 4.5: Summary of the number of intervals.

m'	Minimum	1 ^o Quartile	Median	Mean	Mode	3 ^o Quartile	Maximum
10	4	5	6	6.252	6	7	10
20	6	10	11	11.28	11	12	17
30	8	14	16	16.22	16	18	25
40	7	18.75	21	21.09	21	24	33

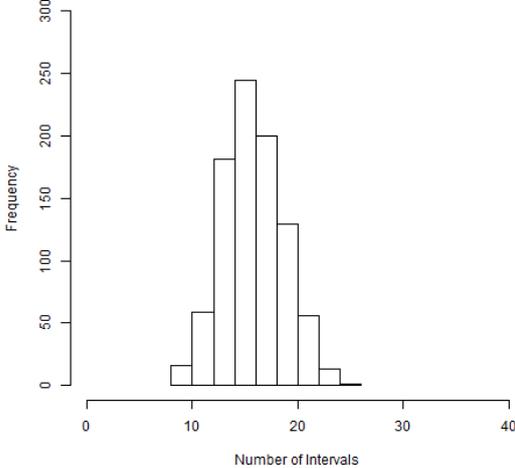
In the same way as the first application, as the number of intervals increases, the discount factor also increases due to the diminishing of data in each interval. This



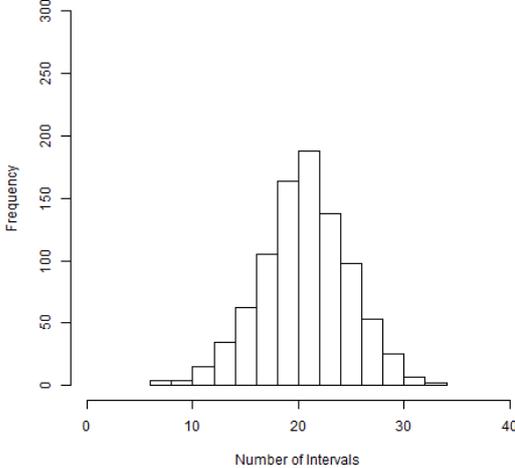
(a) 10 intervals.



(b) 20 intervals.



(c) 30 intervals.



(d) 40 intervals.

Figure 4.9: Histogram of the number of intervals varying the maximum number of intervals.

can be seen in Figure 4.10. It also occurs that the discount factor for the fixed time grid is higher than for the random time grid.

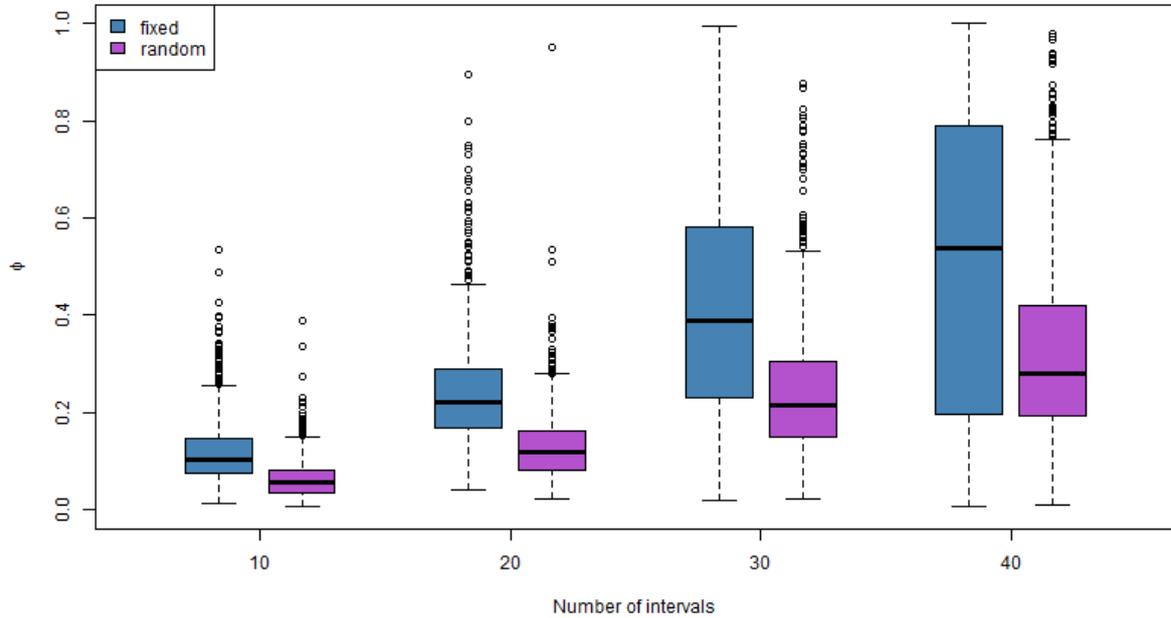


Figure 4.10: Comparison of the discount factor estimates varying the number of intervals.

Table 4.6 contains the Bayes estimates as well as the 95% HPD intervals for each coefficient (for the non-cured individuals and the cure probability) for the chosen model (random time grid with $m' = 20$). Regarding the non-cured individuals, note that the HPD intervals for all the coefficients contain the value 0, it means that the covariates associated with these coefficients does not explain the time until death from melanoma. In turn, to evaluate the cure probability, the most appropriate covariates are age and sex.

Considering that covariate age was standardized and the behavior of the expression representing the probability of cure, $\exp\{-\exp\{\cdot\}\}$, on the domain $(-\infty; \infty)$, the age values that are above the mean (positive values) decrease the probability of cure and, on the contrary, the age values below the mean (negative values), increase this probability. Thus one may have the intuitive conclusion that the older the individual is, the lower is her/his probability of cure. This can be seen in Figure 4.11a.

In turn, the signal of the coefficient related to the covariate sex is negative. This means that the probability of cure will increase for the category represented by 1, which is composed by female individuals. Consequently this probability is lower for the male individuals. This can be seen by Figure 4.11b. This result is also very intuitive

Table 4.6: Estimates of the coefficients associated to the non-cured individuals and the cure fraction.

Non-cured individuals				
	Median	Mean	Standard Deviation	HPD 95%
Age	-0.1477	-0.1463	0.1208	[-0.3913; 0.0674]
Sex (female)	-0.0123	-0.0154	0.1663	[-0.3296; 0.3042]
PS (other)	0.2575	0.2677	0.2763	[-0.2298; 0.8126]

Cure fraction				
	Median	Mean	Standard Deviation	HPD 95%
Intercept	0.3287	0.6592	0.6759	[-0.0098; 2.2392]
Age	0.2116	0.2109	0.0531	[0.0945; 0.3018]
Sex (female)	-0.3851	-0.3858	0.1079	[-0.5838; -0.1610]
PS (other)	0.2523	0.2496	0.1521	[-0.0207; 0.5658]

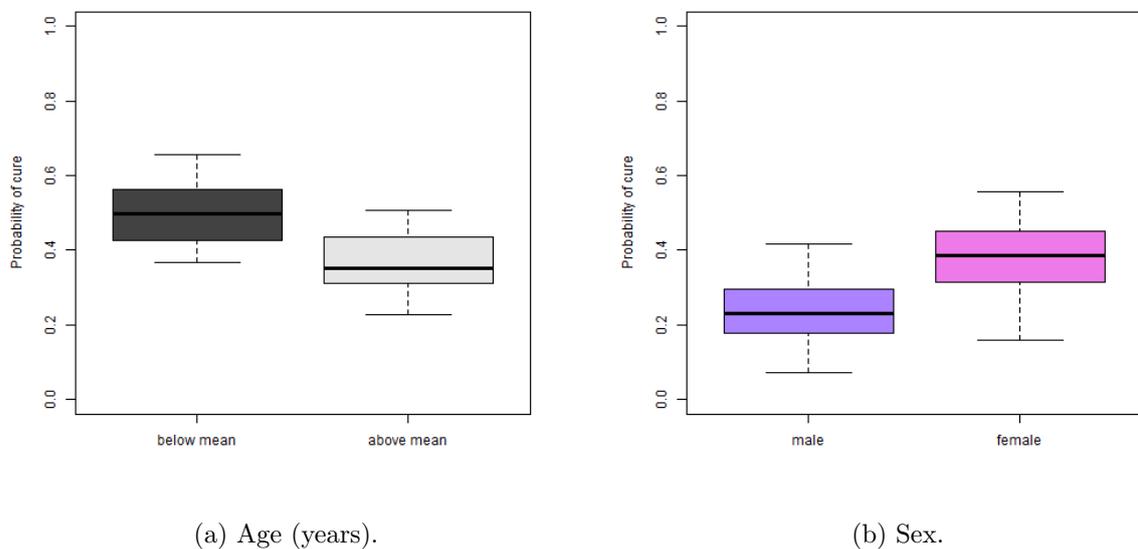


Figure 4.11: Comparison of the probability of cure separated by the covariates age and sex.

when one considers that women generally take better care of the skin and go to the dermatologist more frequently than men, which may lead the discovery of cancer at an early stage.

Another interesting information that can be obtained by estimating the time grid is the most probable grid. According to the chosen model it is $\tau = \{0, 0.4763860, 0.6680356, 0.7939767, 1.5331964, 3.3237509, 4.8815880, 5.6290212, 11.0773443\}$ with probability equal to 0.0033. This grid is made by 8 intervals. Furthermore the probability that the number of intervals is between 8 and 14 is 0.95. Other estimates related to the number of intervals can be seen in Table 4.5. Note that the modal value is 11, which reinforces the choice of the random time grid with $m' = 20$. In turn, Figure 4.9 shows the histogram of the posterior distribution of the number of intervals. Moreover, the median value for the discount factor is 0.1172 and its HPD interval is given by $[0.0286; 0.2694]$.

Lastly, Figure 4.12a shows the estimated population (and improper) survival function (S_{pop}) separated by sex while Figure 4.12b shows the survival function for the non-cured individuals (S_{NC}), also separated by sex. Note that, differently from S_{pop} , the survival function for the non-cured individuals is tending to 0, as the time increases; and S_{pop} stabilizes at a value greater than 0.

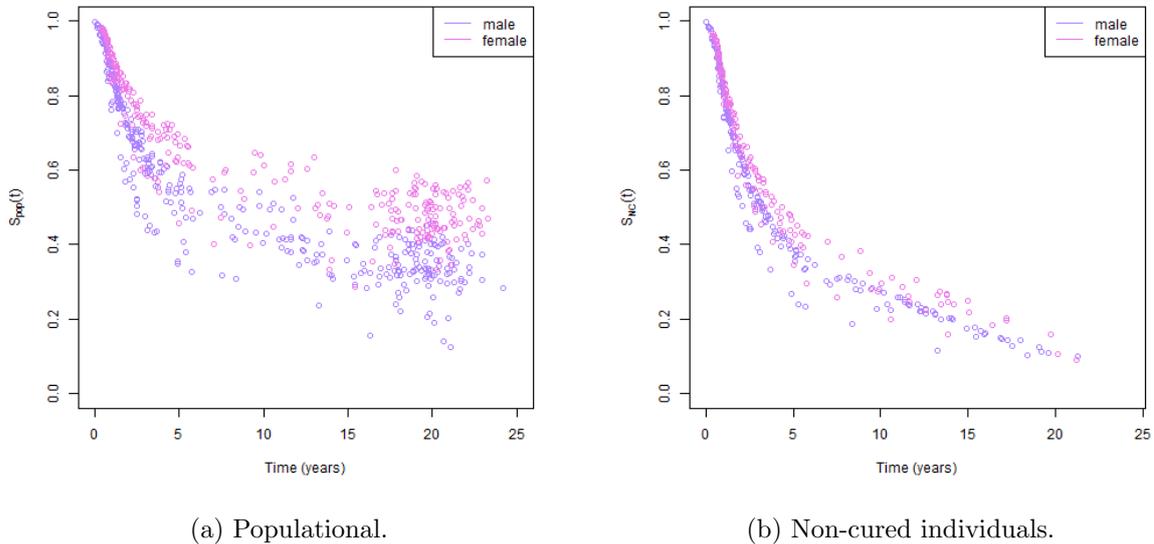


Figure 4.12: Estimated survival functions.

Demarqui et al. (2014) and Kim et al. (2007) analyzed this data set modeling only the cure fraction. In both works, the estimates of the cure fraction coefficients were based on the posterior means and, except by the intercept (ψ_1), the estimates were quite similar to those obtained by the proposed approach. The intercept obtained by Kim et al. (2007) was considerably lower (0.14 for random time grid and other specifications

and 0.136 for fixed time grid and different specifications). The estimates regarding the number of intervals were only similar when comparing to Demarqui et al.'s work, the median value is the exact same, 11, and the mean was 11.103, while on the proposed work was 11.28. Another comparable result is the hazard function which also resulted similarly to the work of Demarqui et al..

5 Conclusions and Future Works

In this master thesis the Piecewise Exponential Model (PEM) was used in the Bayesian context. The parameters of this model, that is, the vector of failure rate was estimated under a dynamic approach, considering a different form of correlate the components of the failure rate. This form enables the achievement of an exact smoothed distribution for these parameters in a simple manner. In the works found in the literature, this distribution is obtained in a approximated way. Therefore, this is a great advantage of the proposed work. Moreover, the time grid of the PEM was estimated via Product Partition Model (PPM) and this whole structure was applied to cure fraction models.

Another advantage of this work is the estimation of the discount factor, which eliminates the necessity of a sensibility study and also allows a complete inference about this quantity. Furthermore, it was observed that the discount factor has the capacity to adapt very well: it gets higher when more information is necessary and lower when there is a good amount of information in the intervals to estimate the parameters (λ).

Turning attention to the time grid, the case in which the criteria of goodness of fit indicated for the fixed time grid, the case of a random grid was immediately after. Thus, one might think that it would be better to lose by not choosing the best model (according to the measures) and, at the same time, to gain by considering the information that can only be obtained when estimating the time grid. Moreover, it was noted that to consider the number of intervals as the number of distinct observed failures is often unnecessary.

Nevertheless, it was observed, in the cure fraction model context, that, as the maximum number of intervals increases, the estimates of the failure rate become dissonant and the intercept of the cure fraction becomes somewhat unstable. This characteristic did not occur neither when the cure rate was estimated under the fixed time grid, nor in the case of the simple model. It may be related to identifiability issues and it should be investigated.

A possible solution for this issue would be to include covariates only on the cure fraction and to verify the behavior of the model, although it is more interesting and

more complete to include this information in both cases. Or, at least, to model the cure probability and the non-cured individuals using different covariates. Informative priors can also be used to model the intercept, once it is possible to obtain prior information from the literature.

It is noteworthy that, although the results obtained in this work are according to the literature, all the conclusions exposed were based on results obtained by applying the proposed model to a real data set, therefore to confirm these results it would be necessary to consider a simulation study. This will be a future work.

A sensibility analysis for the values of the prior distribution of the failure rate associated to the first interval (α_0 and γ_0) would be necessary to verify how far the final estimates are affected by these values; or if this influence is “absorbed” by the smoothed distribution, mainly by obtaining the exact one. A sensibility study would also be interesting for the time grid, both for fixed and random cases. This would be necessary because, even when the grid is being estimated, the number of possible grids is far too large, hence, this type of study would be useful to investigate if the initial grid (that is, that grid formed by the maximum number of intervals admitted *a priori*) affects the results obtained.

An idea to be possibly developed in order to extend the proposed approach, is to consider a function of the discount factor depending on the range of each interval. This idea is motivated by the fact that it would be important to consider the difference between the intervals and their particularities or, in other words, for some intervals more information would be necessary to obtain a good estimation than others. Another motivation is related to the number of events and individuals at risk at the end of follow-up: a better estimation may be obtained if more information was passed through the intervals at this part. So, maybe a unique number would not be enough to capture these differences, mainly on the last intervals.

Another interesting extension to the proposed approach, would be to consider the case of multivariate data.

Lastly, the Mixture Model could be reviewed and a solution to the identifiability issue involving this model could be proposed. For example, to consider informative priors for the intercept, likewise the Promotion Time Model.

A Appendix

In this Chapter there is two sections. The first one is briefly explaining the model comparison methods LPML and WAIC. The second one, in turn, presents the calculations necessary to generate values of the parameters of interest. The idea was to show how these parameters are related to one another and to expose the final expressions.

Comparison Measures

The LPML measure is based on the Conditional Predictive Ordinate (CPO) (Ibrahim et al., 2001b), which, in turn, is a quantity that measures how a specific observation influences the model. For a specific i – th observation, the CPO statistic is given by:

$$CPO_i = f(t_i | D_{obs}^{(-i)}) = \int_{\Phi} f(t_i | \Phi, D_{obs}^{(-i)}) p(\Phi | D_{obs}^{(-i)}) d\Phi, \quad (\text{A.1})$$

where $D_{obs}^{(-i)}$ is the observed data excluding the i – th observation, Φ is a general vector of parameters to be estimated.

Once it is not possible to calculate expression A.1 analytically, an approximation is given by:

$$\widehat{CPO}_i = S \left\{ \sum_{s=1}^S [f(t_i | \Phi^s, D_{obs})]^{-1} \right\}^{-1},$$

where $s = 1, 2, \dots, S$ represents the index of the posterior sample. Finally, the LPML measure can be obtained via:

$$LPML = \sum_{i=1}^n \log(CPO_i) \quad (\text{A.2})$$

The WAIC criteria can be seen as a substitute to the DIC criteria. This measure is an approximation to the following quantity:

$$\text{elpd} = \text{expected log pointwise predictive density for a new dataset} = \sum_{i=1}^n E_{f_i}[\log(p(\tilde{t} | D_{obs}))],$$

where \tilde{t} represents a new observation and D_{obs} is the observed data.

The mentioned approximation is based on the posterior sample and it is given by:

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \hat{p}_{\text{waic}}, \quad (\text{A.3})$$

where

$$\widehat{\text{lpd}} = \text{computed log pointwise predictive density} = \sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S f(t_i | \Phi^s) \right),$$

likewise the notation to the LPML criteria, $s = 1, 2, \dots, S$ is the index associated to the posterior sample, S is the size of the posterior sample and Φ is the vector of estimated parameters.

In turn \hat{p}_{waic} is the estimated the effective number of parameters. It is also estimated using the posterior sample, in the following way:

$$\hat{p}_{\text{waic}} = \sum_{i=1}^n V_{s=1}^S(\log(f(t_i | \Phi^s))), \quad (\text{A.4})$$

where $V_{s=1}^S a_s = \frac{1}{S-1} \sum_{s=1}^S (a_s - \bar{a})^2$ is the sample variance.

Calculations

Basic Model

Considering a random time grid with b intervals, a random sample with size n , no covariates and the following prior distributions $\lambda_1 | D_0 \sim \text{Gamma}(\alpha_0, \gamma_0)$, $(\lambda_j | \phi, D_{j-1}) \sim \text{Gamma}(\phi \alpha_{j-1}, \phi \gamma_{j-1})$, for j in $2, \dots, b$ and $\phi \sim \text{Beta}(\theta_1, \theta_2)$.

The joint posterior distribution of $(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho)$ is given by:

$$\begin{aligned} p(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho | D) &\propto L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho) \propto L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)} | \phi, \rho) p(\phi) p(\rho) \\ &\propto \left[\prod_{j=1}^b \prod_{i=1}^n \lambda_j^{\delta_{ij}} \exp \{-\lambda_j (t_{ij} - s_{j-1})\} \right] \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_1^{\alpha_0-1} \exp \{-\lambda_1 \gamma_0\} \\ &\quad \frac{(\phi \gamma_1)^{\phi \alpha_1}}{\Gamma(\phi \alpha_1)} \lambda_2^{\phi \alpha_1-1} \exp \{-\lambda_2 \phi \gamma_1\} \dots \frac{(\phi \gamma_{b-1})^{\phi \alpha_{b-1}}}{\Gamma(\phi \alpha_{b-1})} \lambda_b^{\phi \alpha_{b-1}-1} \exp \{-\lambda_b \phi \gamma_{b-1}\} \\ &\quad \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \end{aligned} \quad (\text{A.5})$$

The full conditional distribution of $\boldsymbol{\lambda}_{(\rho)}$ is:

$$\begin{aligned}
p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho, D) &\propto L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) \\
&\propto \left[\prod_{j=1}^b \prod_{i=1}^n \lambda_{(\rho)_j}^{\delta_{ij}} \exp\{-\lambda_{(\rho)_j}(t_{ij} - s_{j-1})\} \right] \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_{(\rho)_1}^{\alpha_0-1} \exp\{-\lambda_{(\rho)_1} \gamma_0\} \\
&\quad \frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \lambda_{(\rho)_2}^{\phi\alpha_1-1} \exp\{-\lambda_{(\rho)_2} \phi\gamma_1\} \dots \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \lambda_{(\rho)_b}^{\phi\alpha_{b-1}-1} \exp\{-\lambda_{(\rho)_b} \phi\gamma_{b-1}\} \\
&\quad \text{Replacing } \sum_{i=1}^n \delta_{ij} \text{ by } \eta_j \text{ and } \sum_{i=1}^n (t_{ij} - s_{j-1}) \text{ by } \xi_j: \\
&\propto \lambda_{(\rho)_1}^{\eta_1+\alpha_0-1} \exp\{-\lambda_{(\rho)_1}(\gamma_0 + \xi_1)\} \lambda_{(\rho)_2}^{\eta_2+\phi\alpha_1-1} \exp\{-\lambda_{(\rho)_2}(\phi\gamma_1 + \xi_2)\} \dots \\
&\quad \lambda_{(\rho)_b}^{\eta_b+\phi\alpha_{b-1}-1} \exp\{-\lambda_{(\rho)_b}(\phi\gamma_{b-1} + \xi_b)\} \tag{A.6}
\end{aligned}$$

That is, $(\lambda_{(\rho)_1}|\phi, \rho, D) \sim \text{Gamma}(\eta_1+\alpha_0, \gamma_0+\xi_1)$ and $(\lambda_{(\rho)_j}|\phi, \rho, D) \sim \text{Gamma}(\eta_j+\phi\alpha_{j-1}, \phi\gamma_{j-1} + \xi_j)$ for $j = 2, \dots, b$.

The distribution of $(\phi|\rho, D)$ is given by:

$$\begin{aligned}
p(\phi|\rho, D) &= \int_{\boldsymbol{\lambda}_{(\rho)}} p(\boldsymbol{\lambda}_{(\rho)}, \phi|\rho, D) d\boldsymbol{\lambda}_{(\rho)} \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) p(\phi) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \left[\int_{\lambda_{(\rho)_b}} \dots \int_{\lambda_{(\rho)_2}} \int_{\lambda_{(\rho)_1}} \lambda_{(\rho)_1}^{\sum_{i=1}^n \delta_{i1} + \alpha_0 - 1} \exp\left\{-\lambda_{(\rho)_1} \left(\gamma_0 + \sum_{i=1}^n (t_{i1} - s_0)\right)\right\} \right. \\
&\quad \lambda_{(\rho)_2}^{\sum_{i=1}^n \delta_{i2} + \phi\alpha_1 - 1} \exp\left\{-\lambda_{(\rho)_2} \left(\phi\gamma_1 + \sum_{i=1}^n (t_{i2} - s_1)\right)\right\} \dots \\
&\quad \left. \lambda_{(\rho)_b}^{\sum_{i=1}^n \delta_{ib} + \phi\alpha_{b-1} - 1} \exp\left\{-\lambda_{(\rho)_b} \left(\phi\gamma_{b-1} + \sum_{i=1}^n (t_{ib} - s_{b-1})\right)\right\} \right] \\
&\quad \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \left(\frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \dots \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \right) d\lambda_{(\rho)_1} d\lambda_{(\rho)_2} \dots d\lambda_{(\rho)_b} \\
&\propto \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\eta_j + \phi\alpha_{j-1})}{(\phi\gamma_{j-1} + \xi_j)^{\eta_j + \phi\alpha_{j-1}}} \tag{A.7}
\end{aligned}$$

The distribution of $(\rho|\phi, D)$ is:

$$\begin{aligned}
p(\rho|\phi, D) &= \int_{\boldsymbol{\lambda}_{(\rho)}} p(\boldsymbol{\lambda}_{(\rho)}, \rho|\phi, D) d\boldsymbol{\lambda}_{(\rho)} \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) p(\phi) p(\rho) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \frac{(\gamma_0)^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_0 + \eta_1)}{(\gamma_0 + \xi_1)^{\alpha_0 + \eta_1}} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\phi\alpha_{j-1} + \eta_j)}{(\phi\gamma_{j-1} + \xi_j)^{(\phi\alpha_{j-1} + \eta_j)}} \\
&\propto \frac{(\gamma_0)^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_1)}{(\gamma_1)^{\alpha_1}} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\alpha_j)}{(\gamma_j)^{\alpha_j}} \tag{A.8}
\end{aligned}$$

Including Covariates

In order to include covariates into the model, consider k explanatory variables. The prior distributions for the coefficients associated to the covariates will be $Normal(0, \sigma_l)$, $l = 1, 2, \dots, k$. Given this it was possible to obtain the following expressions.

The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho)$ is:

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho | D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho) \\
&\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\beta}) p(\boldsymbol{\lambda}_{(\rho)} | \phi, \rho) p(\phi) p(\rho) \\
&\propto \left[\prod_{j=1}^b \prod_{i=1}^n (\lambda_{(\rho)_j} \exp\{\mathbf{x}_i \boldsymbol{\beta}\})^{\delta_{ij}} \exp\{-\lambda_{(\rho)_j} \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{ij} - s_{j-1})\} \right] \\
&\quad \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2} \beta_1^2\right\} \dots \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2} \beta_k^2\right\} \\
&\quad \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_{(\rho)_1}^{\alpha_0-1} \exp\{-\lambda_{(\rho)_1} \gamma_0\} \frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \lambda_{(\rho)_2}^{\phi\alpha_1-1} \exp\{-\lambda_{(\rho)_2} \phi\gamma_1\} \dots \\
&\quad \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \lambda_{(\rho)_b}^{\phi\alpha_{b-1}-1} \exp\{-\lambda_{(\rho)_b} \phi\gamma_{b-1}\} \\
&\quad \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \phi^{\theta_1-1} (1 - \phi)^{\theta_2-1} \tag{A.9}
\end{aligned}$$

The full conditional distribution of β_l , for $l = 1, \dots, k$:

$$\begin{aligned}
p(\beta_l | \boldsymbol{\beta}_{(-l)}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho, D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\beta_l) \\
&\propto \left[\prod_{j=1}^b \prod_{i=1}^n (\lambda_{(\rho)_j} \exp\{\mathbf{x}_i \boldsymbol{\beta}\})^{\delta_{ij}} \exp\{-\lambda_{(\rho)_j} \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{ij} - s_{j-1})\} \right] \\
&\quad \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left\{-\frac{1}{2\sigma_l^2} \beta_l^2\right\} \tag{A.10} \\
&\propto \left[\prod_{i=1}^n \exp\left\{\sum_{j=1}^b \delta_{ij} \mathbf{x}_i \beta_l\right\} \right. \\
&\quad \left. \exp\left\{\exp\{\mathbf{x}_i \boldsymbol{\beta}\} \left[-\sum_{j=1}^b \lambda_{(\rho)_j} (t_{ij} - s_{j-1})\right]\right\} \right] \exp\left\{-\frac{1}{2\sigma_l^2} \beta_l^2\right\}
\end{aligned}$$

The full conditional distribution of $\boldsymbol{\lambda}_{(\rho)}$ is:

$$\begin{aligned}
p(\boldsymbol{\lambda}_{(\rho)}|\boldsymbol{\beta}, \phi, \rho, D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D)p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) \\
&\propto \exp\left\{\sum_{i=1}^n \sum_{j=1}^b \delta_{ij}(\mathbf{x}_i\boldsymbol{\beta})\right\} \prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp\left\{-\lambda_{(\rho)_j} \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{ij}-s_{j-1})\right\} \\
&\quad \frac{\gamma_0^{a_0}}{\Gamma(a_0)} \lambda_{(\rho)_1}^{a_0-1} \exp\{-\lambda_{(\rho)_1}\gamma_0\} \frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \lambda_{(\rho)_2}^{\phi\alpha_1-1} \exp\{-\lambda_{(\rho)_2}\phi\gamma_1\} \dots \\
&\quad \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \lambda_{(\rho)_b}^{\phi\alpha_{b-1}-1} \exp\{-\lambda_{(\rho)_b}\phi\gamma_{b-1}\} \\
&\propto \lambda_{(\rho)_1}^{\eta_1+\alpha_0-1} \exp\left\{-\lambda_{(\rho)_1} \left[\gamma_0 + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{i1}-s_0)\right]\right\} \\
&\quad \lambda_{(\rho)_2}^{\eta_2+\phi\alpha_1-1} \exp\left\{-\lambda_{(\rho)_2} \left[\phi\gamma_1 + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{i2}-s_1)\right]\right\} \dots \\
&\quad \lambda_{(\rho)_b}^{\eta_b+\phi\alpha_{b-1}-1} \exp\left\{-\lambda_{(\rho)_b} \left[\phi\gamma_{b-1} + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{ib}-s_{b-1})\right]\right\} \quad (\text{A.11})
\end{aligned}$$

Thereby, $(\lambda_{(\rho)_1}|\boldsymbol{\beta}, \phi, D) \sim \text{Gamma}(\eta_1 + \alpha_0; \gamma_0 + \xi_1)$ and $(\lambda_{(\rho)_j}|\boldsymbol{\beta}, \phi, D) \sim \text{Gamma}(\eta_j + \phi\alpha_{j-1}; \phi\gamma_{j-1} + \xi_j)$, where $\eta_j = \sum_{i=1}^n \delta_{ij}$ and $\xi_j = \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{ij}-s_{j-1})$, for $j = 2, \dots, b$.

The full conditional distribution of ϕ :

$$\begin{aligned}
p(\phi|\boldsymbol{\beta}, \rho, D) &\propto \int_{\boldsymbol{\lambda}_{(\rho)}} p(\boldsymbol{\lambda}_{(\rho)}, \phi|\boldsymbol{\beta}, \rho, D) d\boldsymbol{\lambda}_{(\rho)} \propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho; D) p(\boldsymbol{\lambda}_{(\rho)}|\phi, \rho) p(\phi) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \left[\int_{\lambda_{(\rho)_b}} \dots \int_{\lambda_{(\rho)_2}} \int_{\lambda_{(\rho)_1}} \lambda_{(\rho)_1}^{\eta_1+\alpha_0-1} \exp\left\{-\lambda_{(\rho)_1} \left(\gamma_0 + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{i1}-s_0)\right)\right\} \right. \\
&\quad \lambda_{(\rho)_2}^{\eta_2+\phi\alpha_1-1} \exp\left\{-\lambda_{(\rho)_2} \left(\phi\gamma_1 + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{i2}-s_1)\right)\right\} \dots \\
&\quad \left. \lambda_{(\rho)_b}^{\eta_b+\phi\alpha_{b-1}-1} \exp\left\{-\lambda_{(\rho)_b} \left(\phi\gamma_{b-1} + \sum_{i=1}^n \exp\{\mathbf{x}_i\boldsymbol{\beta}\}(t_{ib}-s_{b-1})\right)\right\} d\lambda_{(\rho)_1} d\lambda_{(\rho)_2} \dots d\lambda_{(\rho)_b} \right] \\
&\quad \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \left(\frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \dots \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \right) \\
&\propto \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\eta_j + \phi\alpha_{j-1})}{(\phi\gamma_{j-1} + \xi_j)^{\eta_j + \phi\alpha_{j-1}}} \\
&\propto \phi^{\theta_1-1} (1-\phi)^{\theta_2-1} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\alpha_j)}{(\gamma_j)^{\alpha_j}} \quad (\text{A.12})
\end{aligned}$$

where ξ_j , for $j = 1, 2, \dots, b$ is defined immediately above.

The distribution of $(\rho|\boldsymbol{\beta}, \phi, D)$ is:

$$\begin{aligned}
p(\rho|\boldsymbol{\beta}, \phi, D) &\propto \int_{\boldsymbol{\lambda}(\rho)} p(\boldsymbol{\lambda}(\rho), \rho|\boldsymbol{\beta}, \phi, D) d\boldsymbol{\lambda}(\rho) \propto \int_{\boldsymbol{\lambda}(\rho)} L(\boldsymbol{\beta}, \boldsymbol{\lambda}(\rho), \phi, \rho; D) p(\boldsymbol{\lambda}(\rho)|\phi, \rho) p(\rho) d\boldsymbol{\lambda}(\rho) \\
&\propto \left[\int_{\lambda_{(\rho)_b}} \dots \int_{\lambda_{(\rho)_2}} \int_{\lambda_{(\rho)_1}} \lambda_{(\rho)_1}^{\eta_1 + \alpha_0 - 1} \exp \left\{ -\lambda_{(\rho)_1} \left(\gamma_0 + \sum_{i=1}^n \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{i1} - s_0) \right) \right\} \right. \\
&\quad \lambda_{(\rho)_2}^{\eta_2 + \phi \alpha_1 - 1} \exp \left\{ -\lambda_{(\rho)_2} \left(\phi \gamma_1 + \sum_{i=1}^n \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{i2} - s_1) \right) \right\} \dots \\
&\quad \left. \lambda_{(\rho)_b}^{\eta_b + \phi \alpha_{b-1} - 1} \exp \left\{ -\lambda_{(\rho)_b} \left(\phi \gamma_{b-1} + \sum_{i=1}^n \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ib} - s_{b-1}) \right) \right\} \right] \\
&\propto \frac{(\gamma_0)^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_1)}{(\gamma_1)^{\alpha_1}} \prod_{j=2}^b \frac{(\phi \gamma_{j-1})^{\phi \alpha_{j-1}}}{\Gamma(\phi \alpha_{j-1})} \frac{\Gamma(\alpha_j)}{(\gamma_j)^{\alpha_j}} \tag{A.13}
\end{aligned}$$

Cure Fraction Model

In the case of cure fraction in the population modeled by the promotion time model and considering the prior distribution for $\boldsymbol{\psi}$, as $Normal(0, \sigma_q^2)$, for $q = k+1, \dots, m$, where k is the number of covariates used to model the non-cured individuals and m is the number of covariates used to model the cured fraction, it is possible to obtain the following distributions:

The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}(\rho), \phi, \boldsymbol{\psi}, \rho)$ is given by:

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\lambda}(\rho), \phi, \boldsymbol{\psi}, \rho|D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}(\rho), \boldsymbol{\psi}, \rho; D) p(\boldsymbol{\beta}) p(\boldsymbol{\lambda}(\rho)|\phi, \rho) p(\phi) p(\boldsymbol{\psi}) p(\rho) \\
&\propto \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ -\lambda_{(\rho)_j} \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \right) \\
&\quad \left(\prod_{i=1}^n \frac{\exp \{ N_i \mathbf{z}_i \boldsymbol{\psi} \} \exp \{ -\exp \{ \mathbf{z}_i \boldsymbol{\psi} \} \}}{N_i!} N_i^{\sum_{j=1}^b \delta_{ij}} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \right) \\
&\quad \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ -\frac{1}{2\sigma_1^2} \beta_1^2 \right\} \dots \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \beta_k^2 \right\} \right) \\
&\quad \left(\frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_{(\rho)_1}^{\alpha_0 - 1} \exp \{ -\lambda_{(\rho)_1} \gamma_0 \} \frac{(\phi \gamma_1)^{\phi \alpha_1}}{\Gamma(\phi \alpha_1)} \lambda_{(\rho)_2}^{\phi \alpha_1 - 1} \exp \{ -\lambda_{(\rho)_2} \phi \gamma_1 \} \dots \right. \\
&\quad \left. \frac{(\phi \gamma_{b-1})^{\phi \alpha_{b-1}}}{\Gamma(\phi \alpha_{b-1})} \lambda_{(\rho)_b}^{\phi \alpha_{b-1} - 1} \exp \{ -\lambda_{(\rho)_b} \phi \gamma_{b-1} \} \right) \\
&\quad \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \phi^{\theta_1 - 1} (1 - \phi)^{\theta_2 - 1} \tag{A.14} \\
&\quad \left(\frac{1}{\sqrt{2\pi\sigma_{k+1}^2}} \exp \left\{ -\frac{1}{2\sigma_{k+1}^2} \psi_1^2 \right\} \dots \frac{1}{\sqrt{2\pi\sigma_{k+m}^2}} \exp \left\{ -\frac{1}{2\sigma_{k+m}^2} \psi_m^2 \right\} \right)
\end{aligned}$$

The full conditional distribution for β_l , $l = 1, 2, \dots, k$ is:

$$\begin{aligned}
p(\beta_l | \boldsymbol{\beta}_{(-l)}, \boldsymbol{\lambda}_{(\rho)}, \phi, \boldsymbol{\psi}, \rho, D_{obs}) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D_{obs}) p(\beta_l) \\
&\propto \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \right) \prod_{i=1}^n \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{z}_i \boldsymbol{\psi} \right\} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \\
&\quad \exp \left\{ - \sum_{j=1}^b \delta_{ij} \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \\
&\quad \exp \left\{ - \exp \{ \mathbf{z}_i \boldsymbol{\psi} \} (1 - \right. \\
&\quad \left. \exp \left\{ - \sum_{j=1}^b \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\}) \right\} \\
&\quad \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp \left\{ \frac{-1}{2\sigma_l^2} \beta_l^2 \right\} \\
&\propto \prod_{i=1}^n \left[\exp \left\{ x_{il} \beta_l \sum_{j=1}^b \delta_{ij} \right\} \right. \\
&\quad \left. \exp \left\{ \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} \left(- \sum_{j=1}^b \lambda_{(\rho)_j} \delta_{ij} (t_{ij} - s_{j-1}) \right) \right\} \right] \quad (\text{A.15}) \\
&\quad \exp \left\{ \exp \{ \mathbf{z}_i \boldsymbol{\psi} \} \sum_{j=1}^b \exp \{ - \lambda_{(\rho)_j} \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \} \right\}
\end{aligned}$$

Then, the full conditional distribution of $\boldsymbol{\lambda}_{(\rho)}$ is given by:

$$\begin{aligned}
p(\boldsymbol{\lambda}_{(\rho)} | \boldsymbol{\beta}, \phi, \boldsymbol{\psi}, \rho, D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) p(\boldsymbol{\lambda}_{(\rho)} | \phi, \rho) \\
&\propto \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ - \lambda_{(\rho)_j} \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ij} - s_{j-1}) \right\} \right) \\
&\quad \left(\prod_{i=1}^n \frac{\exp \{ N_i \mathbf{z}_i \boldsymbol{\psi} \} \exp \{ - \exp \{ \mathbf{z}_i \boldsymbol{\psi} \} \}}{N_i!} N_i^{\sum_{j=1}^b \delta_{ij}} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \right) \\
&\quad \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_{(\rho)_1}^{\alpha_0-1} \exp \{ - \lambda_{(\rho)_1} \gamma_0 \} \frac{(\phi \gamma_1)^{\phi \alpha_1}}{\Gamma(\phi \alpha_1)} \lambda_{(\rho)_2}^{\phi \alpha_1-1} \exp \{ - \lambda_{(\rho)_2} \phi \gamma_1 \} \dots \\
&\quad \frac{(\phi \gamma_{b-1})^{\phi \alpha_{b-1}}}{\Gamma(\phi \alpha_{b-1})} \lambda_{(\rho)_b}^{\phi \alpha_{b-1}-1} \exp \{ - \lambda_{(\rho)_b} \phi \gamma_{b-1} \} \quad (\text{A.16}) \\
&\propto \lambda_{(\rho)_1}^{\alpha_0 + \sum_{i=1}^n \delta_{i1} - 1} \exp \left\{ - \lambda_{(\rho)_1} \left[\gamma_0 + \sum_{i=1}^n N_i \exp \{ x_{i1} \boldsymbol{\beta} \} (t_{i1} - s_0) \right] \right\} \\
&\quad \lambda_{(\rho)_2}^{\phi \alpha_1 + \sum_{i=1}^n \delta_{i2} - 1} \exp \left\{ - \lambda_{(\rho)_2} \left[\phi \gamma_1 + \sum_{i=1}^n N_i \exp \{ x_{i2} \boldsymbol{\beta} \} (t_{i2} - s_1) \right] \right\} \\
&\quad \vdots \\
&\quad \lambda_{(\rho)_b}^{\phi \alpha_{b-1} + \sum_{i=1}^n \delta_{ib} - 1} \exp \left\{ - \lambda_{(\rho)_b} \left[\phi \gamma_{b-1} + \sum_{i=1}^n N_i \exp \{ x_{ib} \boldsymbol{\beta} \} (t_{ib} - s_{b-1}) \right] \right\}
\end{aligned}$$

Thus, $(\lambda_{(\rho)_1} | \boldsymbol{\beta}, \phi, \boldsymbol{\psi}, \rho, D) \sim \text{Gamma}(\alpha_0 + \eta_1, \gamma_0 + \xi_1)$ and $(\lambda_{(\rho)_j} | \boldsymbol{\beta}, \phi, \boldsymbol{\psi}, \rho, D) \sim \text{Gamma}(\phi\alpha_{j-1} + \eta_j, \phi\gamma_{j-1} + \xi_j)$, for $j = 2, \dots, b$, where $\eta_j = \sum_{i=1}^n \delta_{ij}$ and $\xi_j = \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{ij} - s_{j-1})$;

The full conditional distribution of ϕ is:

$$\begin{aligned}
p(\phi | \boldsymbol{\beta}, \boldsymbol{\psi}, \rho, D) &= \int_{\boldsymbol{\lambda}_{(\rho)}} p(\boldsymbol{\lambda}_{(\rho)}, \phi | \boldsymbol{\beta}, \boldsymbol{\psi}, \rho, D) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \int_{\boldsymbol{\lambda}_{(\rho)}} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) p(\boldsymbol{\lambda}_{(\rho)} | \phi, \rho) p(\phi) d\boldsymbol{\lambda}_{(\rho)} \\
&\propto \left[\int_{\lambda_{(\rho)_b}} \cdots \int_{\lambda_{(\rho)_2}} \int_{\lambda_{(\rho)_1}} \right. \\
&\quad \lambda_{(\rho)_1}^{\eta_1 + \alpha_0 - 1} \exp \left\{ -\lambda_{(\rho)_1} \left(\gamma_0 + \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{i1} - s_0) \right) \right\} \\
&\quad \lambda_{(\rho)_2}^{\eta_2 + \phi\alpha_1 - 1} \exp \left\{ -\lambda_{(\rho)_2} \left(\phi\gamma_1 + \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{i2} - s_1) \right) \right\} \cdots \\
&\quad \lambda_{(\rho)_b}^{\eta_b + \phi\alpha_{b-1} - 1} \exp \left\{ -\lambda_{(\rho)_b} \left(\phi\gamma_{b-1} + \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{ib} - s_{b-1}) \right) \right\} d\lambda_{(\rho)_1} \\
&\quad d\lambda_{(\rho)_2} \cdots d\lambda_{(\rho)_b} \left. \right] \\
&\quad \phi^{\theta_1 - 1} (1 - \phi)^{\theta_2 - 1} \left(\frac{(\phi\gamma_1)^{\phi\alpha_1}}{\Gamma(\phi\alpha_1)} \cdots \frac{(\phi\gamma_{b-1})^{\phi\alpha_{b-1}}}{\Gamma(\phi\alpha_{b-1})} \right) \\
&\propto \phi^{\theta_1 - 1} (1 - \phi)^{\theta_2 - 1} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\eta_j + \phi\alpha_{j-1})}{(\phi\gamma_{j-1} + \xi_j)^{\eta_j + \phi\alpha_{j-1}}} \\
&\propto \phi^{\theta_1 - 1} (1 - \phi)^{\theta_2 - 1} \prod_{j=2}^b \frac{(\phi\gamma_{j-1})^{\phi\alpha_{j-1}}}{\Gamma(\phi\alpha_{j-1})} \frac{\Gamma(\alpha_j)}{(\gamma_j)^{\alpha_j}} \tag{A.17}
\end{aligned}$$

Where ξ_j , for $j = 1, 2, \dots, b$ is different from the case with no cure fraction. It is defined immediately above.

The full conditional distribution of ψ_l , for $l = k + 1, \dots, m$ is:

$$\begin{aligned}
p(\psi_l | \boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \phi, \rho, D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}_{(\rho)}, \boldsymbol{\psi}, \rho; D) p(\psi_l) \\
&\propto \left(\prod_{j=1}^b \lambda_{(\rho)_j}^{\sum_{i=1}^n \delta_{ij}} \exp \left\{ -\lambda_{(\rho)_j} \sum_{i=1}^n N_i \exp\{\mathbf{x}_i \boldsymbol{\beta}\} (t_{ij} - s_{j-1}) \right\} \right) \\
&\quad \left(\prod_{i=1}^n \frac{\exp\{N_i \mathbf{z}_i \boldsymbol{\psi}\} \exp\{-\exp\{\mathbf{z}_i \boldsymbol{\psi}\}\}}{N_i!} N_i^{\sum_{j=1}^b \delta_{ij}} \exp \left\{ \sum_{j=1}^b \delta_{ij} \mathbf{x}_i \boldsymbol{\beta} \right\} \right) \\
&\quad \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp \left\{ \frac{-1}{2\sigma_l^2} \psi_l^2 \right\} \\
&\propto \exp \left\{ \sum_{i=1}^n (N_i \mathbf{z}_i \boldsymbol{\psi}) - \sum_{i=1}^n \exp\{\mathbf{z}_i \boldsymbol{\psi}\} \right\} \exp \left\{ -\frac{1}{2\sigma_l^2} \psi_l^2 \right\} \tag{A.18}
\end{aligned}$$

The distribution of $(\rho|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, D)$ is:

$$\begin{aligned}
p(\rho|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, D) &= \int_{\boldsymbol{\lambda}(\rho)} p(\boldsymbol{\lambda}(\rho), \rho|\boldsymbol{\beta}, \phi, \boldsymbol{\psi}, D) d\boldsymbol{\lambda}(\rho) \\
&\propto \int_{\boldsymbol{\lambda}(\rho)} L(\boldsymbol{\beta}, \boldsymbol{\lambda}(\rho), \boldsymbol{\psi}, \rho; D) p(\boldsymbol{\lambda}(\rho)|\phi, \rho) p(\rho) d\boldsymbol{\lambda}(\rho) \\
&\propto \left[\int_{\lambda_{(\rho)b}} \cdots \int_{\lambda_{(\rho)2}} \int_{\lambda_{(\rho)1}} \right. \\
&\quad \lambda_{(\rho)1}^{\eta_1 + \alpha_0 - 1} \exp \left\{ -\lambda_{(\rho)1} \left(\gamma_0 + \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{i1} - s_0) \right) \right\} \\
&\quad \lambda_{(\rho)2}^{\eta_2 + \phi \alpha_1 - 1} \exp \left\{ -\lambda_{(\rho)2} \left(\phi \gamma_1 + \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{i2} - s_1) \right) \right\} \cdots \\
&\quad \left. \lambda_{(\rho)b}^{\eta_b + \phi \alpha_{b-1} - 1} \exp \left\{ -\lambda_{(\rho)b} \left(\phi \gamma_{b-1} + \sum_{i=1}^n N_i \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} (t_{ib} - s_{b-1}) \right) \right\} \right] \\
&\propto \frac{(\gamma_0)^{\alpha_0} \Gamma(\alpha_1)}{\Gamma(\alpha_0) (\gamma_1)^{\alpha_1}} \prod_{j=2}^b \frac{(\phi \gamma_{j-1})^{\phi \alpha_{j-1}} \Gamma(\alpha_j)}{\Gamma(\phi \alpha_{j-1}) (\gamma_j)^{\alpha_j}} \tag{A.19}
\end{aligned}$$

Bibliography

- Arjas, E. and Gasbarra, D. (1994) Nonparametric bayesian inference from right censored survival data. *Statistica Sinica*, **4**, 505 – 524.
- Banerjee, S. and Carlin, B. P. (2004) Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, **60**, 268 – 275.
- Barreto-Souza, W. (2015) Long-term survival models with overdispersed number of competing causes. *Computational Statistics & Data Analysis*, **91**, 51 – 63.
- Barry, D. and Hartigan, J. A. (1992) Product partition models for change point problems. *The Annals of Statistics*, **20**, 260 – 279.
- Basu, S. and Tiwari, R. C. (2010) Breast cancer survival, competing risks and mixture cure model: a bayesian analysis. *Journal of the Royal Statistical Society*, **173**, 307 – 329.
- Berkson, J. and Gage, R. P. (1952) Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501 – 515.
- Birnbaum, Z. W. and Saunders, S. C. (1969) A new family of life distributions. *Journal of Applied Probability*, **6**, 319 – 327.
- Boag, J. W. (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, **11**, 15 – 53.
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campo, D. P., Barbosa, M. T. S. and Shimakura, S. E. (2011) *Análise de Sobrevivência: teoria e aplicações em saúde*. Editora Fiocruz, 2 edn.
- de Castro, M., Cancho, V. G. and Rodrigues, J. (2009) A bayesian long-term survival model parametrized in the cured fraction. *Biometrical Journal*, **51**, 443 – 455.
- de Castro, M., Chen, M.-H. and Zhang, Y. (2015) Bayesian path specific frailty models for multi-state survival data with applications. *Biometrics*, **71**, 760 – 771.

- Chen, M.-H., Ibrahim, J. G. and Sinha, D. (1999) A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909 – 919.
- Colosimo, E. A. and Giolo, S. R. (2006) *Análise de Sobrevida Aplicada*. ABE - Projeto Fisher. Edagrd Blücher.
- Cox, D. R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187 – 220.
- Cucchetti, A., Ferrero, A., Cescon, M., Donadon, M., Russolillo, N., Ercolani, G., Stacchini, G., Mazzotti, F., Torzilli, G. and Pinna, A. D. (2015) Cure model survival analysis after hepatic resection for colorectal liver metastases. *Annals of Surgical Oncology*, **22**, 1908 – 1914.
- Demarqui, F. N. (2010) *Uma classe mais flexível de modelos semiparamétricos para dados de sobrevivência*. Ph.D. thesis, Universidade Federal de Minas Gerais.
- Demarqui, F. N., Dey, D. K., Loschi, R. H. and Colosimo, E. A. (2014) Fully semiparametric bayesian approach for modeling survival data with cure fraction. *Biometrical Journal*, **56**, 198 – 218.
- Farewell, V. T. (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041 – 1046.
- (1986) Mixture models in survival analysis: Are they worth the risk? *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, **14**, 257 – 262.
- Gamerman, D. (1991) Dynamic bayesian models for survival data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **40**, 63 – 79.
- (1994) Bayes estimation of the piece-wise exponential distribution. *IEEE Transactions on Reliability*, **43**, 128 – 131.
- Gamerman, D. and Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, 2 edn.
- Gamerman, D., dos Santos, T. R. and Franco, G. C. (2013) A non-gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, **35**, 625 – 645.

- Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995) Adaptive rejection metropolis sampling within gibbs sampling. *Applied Statistics*, **44**, 455 – 472.
- Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **41**, 337 – 348.
- Hartigan, J. (1990) Partition models. *Communications in Statistics - Theory and Methods*, **19**, 2745 – 2756.
- Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001a) Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, **57**, 383 – 388.
- (2001b) *Bayesian Survival Analysis*. Springer Series in Statistics. Springer, 1 edn.
- Kalbfleisch, J. D. and Prentice, R. L. (1973) Marginal likelihoods based on cox's regression and life model. *Biometrika*, **60**, 267 – 278.
- Kim, S., Chen, M.-H., Dey, D. K. and Gamerman, D. (2007) Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, **13**, 17 – 35.
- Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G. and Scheike, T. H. (2014) *Handbook of Survival Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.
- Klein, J. P. and Moeschberger, M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer-Verlag New York, 2 edn.
- Lambert, P. C. (2007) Modeling of the cure fraction in survival studies. *The Stata Journal*, **7**, 351 – 375.
- Lambert, P. C. and Thompson, J. R. (2007) Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, **8**, 576 – 594.
- Lawless, J. F. (2002) *Statistical Models and Methods for Lifetime Data*. Wiley - Interscience, 2 edn.
- Li, C. S., Taylor, J. M. G. and Sy, J. P. (2001) Identifiability of cure models. *Statistics and Probability Letters*, **54**, 389 – 395.

- Loschi, R. and Cruz, F. (2005) Extension to the product partition model: computing the probability of a change. *Computational Statistics & Data Analysis*, **48**, 255 – 268.
- McKeague, I. W. and Tighiouart, M. (2000) Bayesian estimators for conditional hazard functions. *Biometrics*, **56**, 1007 – 1015.
- Peng, Y. (2003) Estimating baseline distribution in proportional hazards cure models. *Computational Statistics & Data Analysis*, **42**, 187 – 201.
- Petris, G. (2010) An R package for dynamic linear models. *Journal of Statistical Software*, **36**, 1 – 16.
- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rodrigues, J., Cancho, V. G. and de Castro, M. (2008) Teoria unificada de análise de sobrevivência. *Tech. rep.*, Universidade Federal de São Carlos, Universidade de São Paulo.
- Rodrigues, J., Cancho, V. G., de Castro, M. and Louzada-Neto, F. (2009) On the unification of long-term survival models. *Statistics & Probability Letters*, **79**, 753 – 759.
- Sinha, D., Chen, M.-H. and Ibrahim, J. G. (2003) Bayesian inference for survival data with a surviving fraction. *Lecture Notes-Monograph Series*, **43**, 117 – 138.
- Taylor, J. M. G. (1995) Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899 – 907.
- Vehtari, A. and Gelman, A. (2014) WAIC and cross-validation in Stan. -.
- Watanabe, S. (2010) Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**, 3571 – 3594.
- West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer, 2 edn.
- Yakovlev, A. Y. and Tsodikov, A. D. (1996) *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific.

-
- Yin, G. and Ibrahim, J. G. (2005) A general class of bayesian survival models with zero and nonzero cure fractions. *Biometrics*, **61**, 403 – 412.
- Yu, B., Tiwari, R. C., Cronin, K. A. and Feuer, E. J. (2004) Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, **23**, 1733 – 1747.