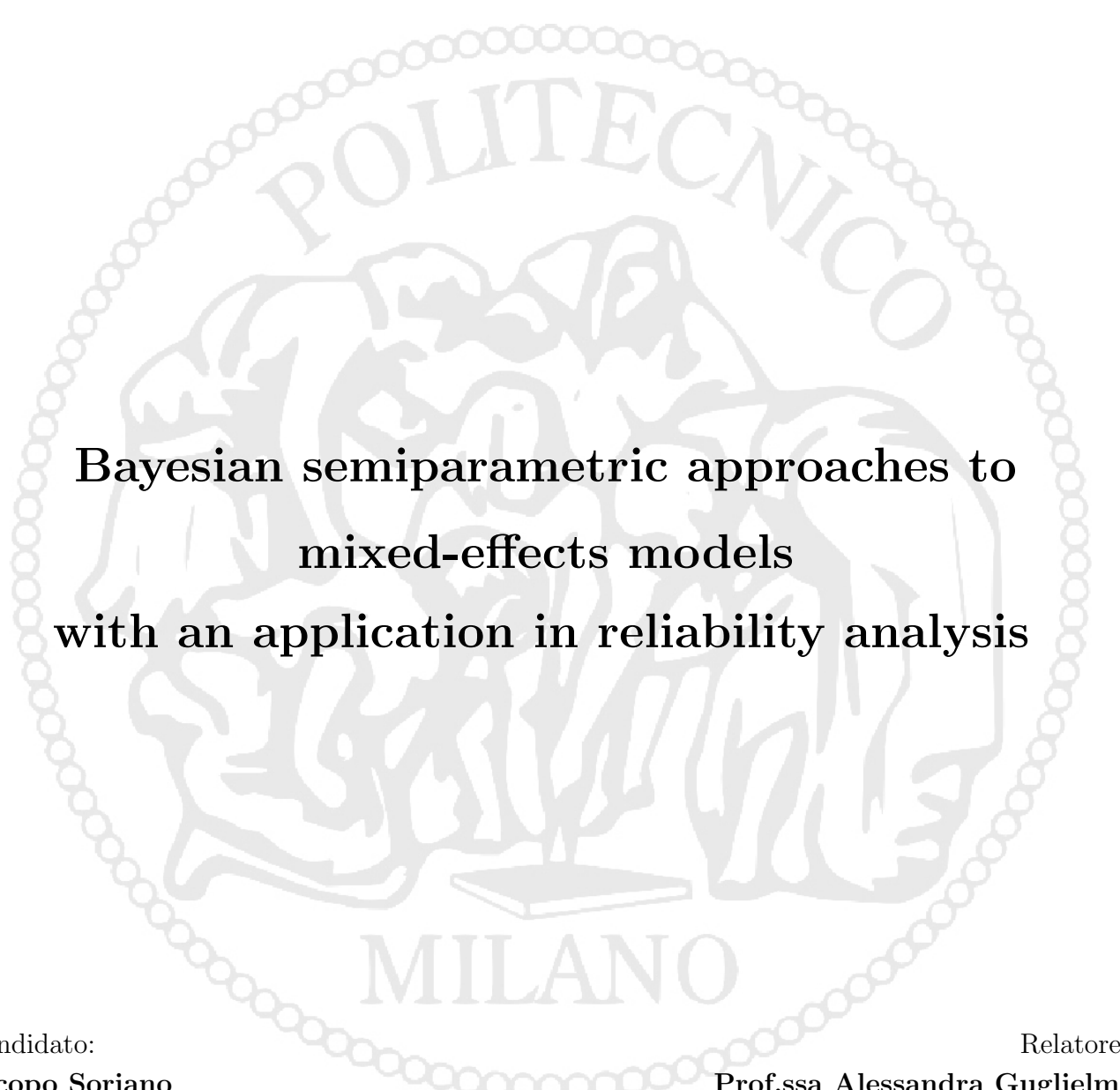


**Politecnico di Milano**

---

FACOLTÀ DI INGEGNERIA DEI SISTEMI  
Corso di Laurea in Ingegneria Matematica

TESI DI LAUREA SPECIALISTICA



**Bayesian semiparametric approaches to  
mixed-effects models  
with an application in reliability analysis**

Candidato:  
**Jacopo Soriano**  
Matricola 735664

Relatore:  
**Prof.ssa Alessandra Guglielmi**

Correlatore:  
**Dott. Raffaele Argiento**

---

Anno Accademico 2009–2010

*“Prediction is very difficult, especially about the future.”*

Niels Bohr

## **Ringraziamenti**

Molte persone hanno permesso, in modo più o meno diretto, che io arrivassi a realizzare questa tesi e questo è il luogo per esprimere loro tutta la mia gratitudine.

Un primo ringraziamento va alla professoressa Alessandra Guglielmi per avermi guidato con entusiasmo e attenzione durante tutto il lavoro della tesi, e per aver dimostrato grande fiducia nelle mie capacità e nella mia persona. Ringrazio molto anche il dottor Raffaele Argiento per la totale disponibilità e per il buon umore con cui ha risposto a tutte le mie domande.

Un grazie alla mia famiglia che è sempre stata un punto di riferimento: mi ha sostenuto nelle situazioni difficili e ha gioito con me in quelle felici. Un grazie a mia nonna Dina che mi ha trasmesso la passione per la matematica sin dalle tabelline.

Un grazie a tutti gli amici del Liceo, del Politecnico e di Centrale con i quali ho potuto confrontarmi e crescere tra e fuori dai banchi.

Un grazie a Elisa, sempre al mio fianco.

# Contents

<b>Introduzione</b>	<b>1</b>
<b>Introduction</b>	<b>5</b>
<b>1 Bayesian Nonparametrics</b>	<b>9</b>
1.1 Exchangeability assumption . . . . .	9
1.2 Dirichlet processes . . . . .	11
1.3 Normalized random measures with independent increments . . . . .	13
1.4 Normalized generalized gamma processes . . . . .	15
<b>2 Generalized Linear Mixed Models</b>	<b>19</b>
2.1 Linear mixed models . . . . .	20
2.2 Generalized linear mixed models . . . . .	20
2.3 Bayesian approaches to the generalized linear mixed models . . . . .	21
<b>3 Application to accelerated failure time models</b>	<b>23</b>
3.1 Survival analysis . . . . .	23
3.2 Accelerated life test on NASA pressure vessels . . . . .	24
3.2.1 NASA pressure vessels and Kevlar fibres . . . . .	25
3.3 Accelerated life models for Kevlar fiber failure times . . . . .	26
3.3.1 Parametric model . . . . .	27
3.3.2 Nonparametric error model . . . . .	28
3.3.3 Nonparametric random effects model . . . . .	32
3.3.4 DPpackage: gamma-distributed error model . . . . .	34
<b>4 Results</b>	<b>36</b>
4.1 Nonparametric error model . . . . .	38
4.1.1 Interval estimates for a new random spool and for given ones . . . . .	39
4.1.2 Posterior estimates . . . . .	43
4.1.3 Predicted survival functions . . . . .	46

4.1.4	Residuals . . . . .	48
4.1.5	Computation details . . . . .	50
4.2	Nonparametric random effects model . . . . .	52
4.2.1	Interval estimates for a new random spool and for given ones . . . . .	52
4.2.2	Posterior estimates . . . . .	56
4.2.3	Predicted survival functions . . . . .	58
4.2.4	Residuals . . . . .	59
4.2.5	Computation details . . . . .	60
4.3	DPpackage: gamma-distributed error model . . . . .	61
4.3.1	Interval estimates for a new random spool and for given ones . . . . .	62
4.3.2	Posterior estimates . . . . .	64
4.3.3	Computation details . . . . .	65
4.4	Conclusions . . . . .	66
<b>A</b>	<b>Full-conditionals of the nonparametric error model</b>	<b>67</b>
A.1	Parametric full-conditionals . . . . .	67
A.2	Full-conditionals of the error . . . . .	70
<b>B</b>	<b>Full-conditionals of the nonparametric random effects model</b>	<b>73</b>
B.1	Parametric full-conditionals . . . . .	74
B.2	Full-conditionals of the random effects . . . . .	75
<b>C</b>	<b>Data-set</b>	<b>80</b>
	<b>Bibliography</b>	<b>82</b>

# List of Figures

3.1	Photos of the NASA pressure vessels from the 2009 Composite Pressure Vessel and Structure Summit [5]. . . . .	26
4.1	Marginal error distribution density functions for different hyperparameters of the gamma prior for the shape parameter $\theta$ . (a) Marginal error of Leon et al.'s parametric model, and (b) marginal error used in our semiparametric models. . . . .	37
4.2	Prior and posterior number of mixture components for the model with nonparametric error . . . . .	42
4.3	Posterior kernel density estimation of the random effects for the model with nonparametric error ( $\sigma = 0.3$ and $\kappa = 0.3$ ) . . . . .	44
4.4	Posterior marginal density distribution functions of the error in the model with nonparametric error . . . . .	45
4.5	Survival functions for a new random spool at 22.5MPa for the model with nonparametric error . . . . .	46
4.6	Predicted survival functions for a new random spool at 23.4MPa for the model with nonparametric error ( $\sigma = 0.3$ , $\kappa = 0.3$ ), and for Kaplan-Meier estimator . . . . .	47
4.7	Bayesian residuals for the uncensored failure times under the model with nonparametric error ( $\sigma = 0.3$ , $\kappa = 0.3$ ). . . . .	49
4.8	Markov chain sample of $\beta_0$ , $\beta_1$ and $\alpha_1$ . . . . .	50
4.9	Scatterplot of the log-survival times against the log-stress for the two semiparametric models. . . . .	53
4.10	Posterior kernel density estimation of the random effects for the model with nonparametric random effects ( $\sigma = 0.3$ and $\kappa = 1.2$ ) . . . . .	57
4.11	(a) Posterior kernel density estimation of the shape parameter $\theta$ and (b) posterior marginal density distribution function of the error for the model with nonparametric error, where $\sigma = 0.3$ and $\kappa = 1.2$ . . . . .	57
4.12	Predicted survival functions for a new random spool for the model with nonparametric random effects ( $\sigma = 0.3$ and $\kappa = 1.2$ ) . . . . .	58

4.13 Bayesian residuals against the fitted log failure times . . . . .	59
4.14 Bayesian residuals against the quantiles of the Gumbel standard at the four stress levels . . . . .	60
4.15 Markov chain sample of $\beta_1$ and $\alpha_1$ . . . . .	61
4.16 Density distribution function of the marginal error $V$ for different choices of the hyperparameters $a_0$ and $b_0$ . . . . .	62
4.17 Markov chain sample of $\beta_1$ and of the number of clusters $K_J$ . . . . .	66

# List of Tables

3.1	Mechanical properties of a single fiber and a single yarn of Kevlar . . . . .	26
4.1	Interval estimates of the quantiles of the predictive distributions for the parametric model . . . . .	38
4.2	Interval estimates of the quantiles of the predictive distributions for the model with nonparametric error for different hyperparameters of the NGG process prior. . . . .	40
4.3	Interval estimates of the quantiles of the predictive distributions for the model with nonparametric error ( $\sigma = 0.3, \kappa = 0.3$ ) . . . . .	41
4.4	Posterior mean and standard deviation (SD) of the effects for the model with nonparametric error ( $\sigma = 0.3$ and $\kappa = 0.3$ ) . . . . .	43
4.5	Interval estimates of the quantiles of the predictive distributions for different number of grid points for the model with nonparametric error ( $\sigma = 0.3, \kappa = 0.1$ ). . . . .	51
4.6	Interval estimates of the quantiles of the predictive distributions for the model with nonparametric random effects ( $\sigma = 0.3, \kappa = 1.2$ ) . . . . .	54
4.7	Interval estimates of the quantiles of the predictive distributions for the model with nonparametric random effects for different hyperparameters of the NGG process prior. . . . .	55
4.8	Posterior mean and standard deviation (SD) of the effects for the model with nonparametric random effects ( $\sigma = 0.3$ and $\kappa = 1.2$ ) . . . . .	56
4.9	Interval estimates of the quantiles of the predictive distributions for the model with gamma error for different hyperparameters of the dispersion parameter $\theta$ ( $a \sim Gamma(a_1 = 1, b_1 = 1)$ ). . . . .	63
4.10	Interval estimates of the quantiles of the predictive distributions for the model with gamma error for different hyperparameters of the mass parameter $a$ ( $\theta \sim Gamma(a_0 = 1, b_0 = 1)$ ). . . . .	63
4.11	Interval estimates of the quantiles of the predictive distributions for the model with gamma error ( $\theta \sim Gamma(a_0 = 1, b_0 = 1)$ and $a \sim Gamma(a_1 = 1, b_1 = 1)$ ) . . . . .	64



4.12	Posterior mean and standard deviation (SD) of the effects for the model with gamma error ( $\theta \sim \text{Gamma}(a_0 = 1, b_0 = 1)$ and $a \sim \text{Gamma}(a_1 = 1, b_1 = 1)$ )	64
C.1	Failure-times of NASA pressure vessels wrapped with Kevlar. Right-censored observations are indicated with an asterisk *.	80

## Abstract

Questo lavoro si prefigge un duplice obiettivo: da una parte quello di effettuare uno studio metodologico sui modelli ad effetti casuali in ambito bayesiano non parametrico, dall'altro di utilizzare questi modelli in un problema di affidabilità. Dopo una breve introduzione all'approccio bayesiano non parametrico, segue la costruzione del processo gamma generalizzato normalizzato (NGG) a partire da misure di probabilità aleatorie a incrementi indipendenti, e una descrizione delle principali proprietà. Tale processo sarà utilizzato come ingrediente nei modelli presentati in seguito. I test di rottura accelerata (ALT) sono test di affidabilità in cui si accelera il processo di rottura di un pezzo meccanico con l'obiettivo di estrapolare quale sarà la vita del componente a condizioni di stress normali. I dati relativi a un ALT possono essere analizzati con un modello accelerated failure time (AFT) in cui si cerca di comprendere il legame tra il logaritmo del tempo di rottura e alcune variabili esplicative. In questo lavoro analizziamo un ALT effettuato dalla NASA su alcuni recipienti in pressione dello Space Shuttle. I recipienti sono rivestiti con Kevlar proviente da diversi rocchetti e trattiamo l'effetto del rocchetto come casuale. I tempi di rottura sono rappresentati con due modelli AFT bayesiani semiparametrici, con l'obiettivo di fornire intervalli di credibilità per determinati quantili della distribuzione del tempo di vita di un recipiente rivestito con Kevlar proveniente da un nuovo rocchetto. Nel primo modello, l'errore è rappresentato da una mistura di distribuzioni parametriche in cui la misturante è un processo NGG. Nel secondo modello, gli effetti casuali vengono considerati in modo non parametrico, utilizzando una prior di tipo NGG. Per entrambi i modelli abbiamo calcolato analiticamente le espressioni delle full-conditional necessarie per poter costruire un algoritmo MCMC che permetta di campionare dalla distribuzione a posteriori; quindi abbiamo implementato gli algoritmi in C ed eseguito simulazioni numeriche. In particolare, nel primo modello ad ogni iterazione dell'algoritmo simuliamo una traiettoria dal processo NGG, mentre nel secondo usiamo un algoritmo di tipo Polya-urn.

## Abstract

In this work we provide a methodological study about Bayesian nonparametric random-effects models, and an application of these models in reliability. After a brief introduction to the nonparametric Bayesian approach, the construction of the normalized generalized gamma process (NGG) by normalization of a completely random measure is provided. This process is an ingredient of the models we will introduce later. Accelerated life testing (ALT) involves acceleration of failure times with the purpose of predicting the life-time of the product at normal use conditions. Data from an ALT can be analyzed by a so-called Accelerated Failure Time (AFT) model, where the dependence between the logarithm of the failure time is related to some explanatory variables. We analyze an AFT made by NASA on some pressure vessels, which are critical components of the Space Shuttle, via two semi-parametric Bayesian AFT models. The pressure vessels are wrapped with Kevlar from different spools and we treated the spool effect as random. In particular, we provide credibility intervals of some given quantiles of the failure-time distribution for a pressure vessel wrapped with fiber from a new random spool. In the first model the error is represented by a mixture of parametric distribution with a NGG mixing measure, while in the second one the random effects have a NGG process prior. For both models, we derived the analytical expressions of the full-conditional distributions needed to make a MCMC algorithm to sample from the posterior distribution; then we coded the algorithms in C and we made numerical simulations. In particular, at each iteration of the first models algorithm, we sample a trajectory of the NGG process; while in the second model, we implemented a Polya-urn scheme algorithm.

# Introduzione

Questo lavoro si prefigge un duplice obiettivo: da una parte quello di effettuare uno studio metodologico sui modelli ad effetti casuali in ambito bayesiano non parametrico, dall'altro di utilizzare questi modelli in un problema di ingegneria dell'affidabilità.

I modelli ad effetti casuali sono modelli che separano la variabilità tra diverse unità statistiche (gruppi) dalla variabilità all'interno di ogni singola unità. Questi modelli vengono anche chiamati modelli gerarchici in quanto permettono appunto di definire due o più livelli di variabilità. Vengono spesso applicati in ambito clinico, ad esempio, dove vengono ripetute diverse osservazioni (misurazioni) per ogni paziente con l'obiettivo di cogliere sia eventuali differenze tra i pazienti sia variazioni a livello del singolo soggetto; in questo caso il fattore di raggruppamento è il paziente. Questi modelli vengono utilizzati anche in altri ambiti, tra i quali citiamo l'analisi di affidabilità di cui ci occuperemo più tardi. Un'ampia classe di modelli ad effetti casuali è quella dei modelli lineari generalizzati a effetti misti (GLMM) in cui gli effetti sono sia fissi che casuali, e in cui si assume che le risposte abbiano distribuzione appartenente alla famiglia esponenziale. Secondo Gelman [13], diremo che un effetto è fisso se è identico per tutti i gruppi, e casuale se varia da gruppo a gruppo.

L'approccio bayesiano agli effetti casuali ha diversi vantaggi rispetto a quello frequentista. Prima di tutto l'inferenza frequentista si basa su stime asintotiche, mentre in ambito bayesiano è sempre possibile fare inferenza esatta anche per data-set di piccola dimensione, utilizzando metodi di integrazione numerica Markov Chain Monte Carlo (MCMC). Inoltre, in statistica classica gli effetti casuali tipicamente sono assunti tra loro indipendenti; mentre in statistica bayesiana vengono assunti scambiabili e questo legame tra gli effetti rende le stime più accurate. In particolare, si ottengono stime ragionevoli anche per quelle unità statistiche, rappresentate dagli effetti casuali, che contengono pochi soggetti. Ricordiamo che l'assunzione di scambiabilità è ragionevole in quanto le unità statistiche possono essere interpretate come un campione senza rimpiazzo dalla popolazione delle unità statistiche.

Spesso non è ragionevole fare specifiche assunzioni sulla distribuzione degli effetti casuali e quindi, in questi casi, l'utilizzo di un modello parametrico porterebbe a stime distorte dei parametri del modello. L'approccio non parametrico rilassa l'ipotesi di appartenenza di una distribuzione ad una classe parametrica, permettendo di superare la dipendenza da

assunzioni parametriche. Inoltre una prior non parametrica risulta più flessibile e permette un'inferenza più robusta.

Il limite dei modelli bayesiani non parametrici è il grande sforzo computazionale richiesto, anche se la crescente disponibilità delle capacità di calcolo degli ultimi anni ha reso molto popolare l'inferenza bayesiana non parametrica e ha portato ad una grossa produzione di modelli bayesiani non parametrici in letteratura. Nonostante ciò l'onere computazionale è tuttora un problema di primaria importanza e quindi la ricerca di algoritmi sempre più efficienti resta attuale. Per una presentazione delle più comuni classi di prior non parametriche e dei principali tecniche di inferenza non parametrica si veda Müller e Quintana [29]. Di recente, l'approccio bayesiano non parametrico ha riscontrato successo soprattutto in biostatistica; si veda Dunson [7] per una descrizione esaustiva dei più recenti modelli bayesiani non parametrici utilizzati in questo ambito. In questo lavoro concentriamo la nostra attenzione sulla prior nonparametrica Normalized Generalized Gamma Process (processo NGG). Questo processo, introdotto da Brix [3] nella versione non normalizzata e studiato per la prima volta come prior da Regazzini et al. [35], si costruisce come normalizzazione di una misura completamente aleatoria e comprende il processo di Dirichlet come caso particolare. Così come il processo di Dirichlet, il processo NGG seleziona distribuzioni discrete con probabilità uno, e induce una partizione aleatoria sugli interi positivi; in particolare, questo vuol dire che se consideriamo un campione aleatorio di ampiezza  $n$  da questo processo, tra le  $n$  realizzazioni ci potranno essere dei valori ripetuti e questi valori ripetuti definiscono una partizione degli interi  $1, \dots, n$ . Questa partizione aleatoria è governata da due parametri nel caso dell'NGG, a differenza del processo di Dirichlet in cui il clustering è governato da un solo parametro. Questo ulteriore grado di libertà rende il clustering più ricco e quindi la prior più flessibile.

I test di rottura accelerata (ALT) hanno grande importanza in affidabilità. Si tratta di test in cui si accelera il processo di rottura di un pezzo meccanico con l'obiettivo di estrapolare quale sarà la vita del componente a condizioni di stress normali. I dati relativi ad un ALT posso essere analizzati con i cosiddetti modelli "accelerated failure time" (AFT) in cui si cerca di comprendere il legame fra il logaritmo del tempo di rottura (o tempo di vita) e alcune variabili esplicative come ad esempio la pressione o la temperatura; si tratta dunque, in scala logaritmica, di un modello di regressione lineare. Spesso è ragionevole assumere che il tempo di rottura abbia distribuzione Weibull: visto che la distribuzione Weibull non appartiene alla famiglia esponenziale, il corrispondente modello AFT non è un modello lineare generalizzato ma una sua naturale estensione. In questo lavoro analizzeremo un ALT effettuato dalla NASA su alcuni recipienti in pressione dello Space-Shuttle. Il data-set consiste di 108 tempi di rottura di questi particolari recipienti in pressione ricoperti in fibra di Kevlar, soggetti a 4 livelli di stress, e con fibra proveniente da 8 diversi rocchetti. Il problema dell'affidabilità di questi recipienti è ancora di attualità, infatti nel

2009 la NASA ha organizzato il *Composite Pressure Vessel and Structure Summit* [5] per discutere i progressi in questo ambito.

In questo lavoro i tempi di rottura vengono modellati attraverso un modello AFT bayesiano semiparametrico, con l'obiettivo di fornire intervalli di credibilità per determinati quantili della distribuzione del tempo di vita di un nuovo recipiente ricoperto con fibra di Kevlar proveniente da un nuovo rocchetto. Dal momento che abbiamo diverse osservazioni per ogni rocchetto, risulta naturale considerare l'effetto del rocchetto come casuale; mentre trattiamo l'effetto della pressione come effetto fisso. Questo data-set è stato studiato da diversi autori con approcci frequentisti, in cui sia rocchetto che pressione venivano trattati come effetti fissi; più recentemente Leon et al. [24] invece hanno proposto un modello AFT bayesiano parametrico ad effetti misti. Riteniamo che le stime ottenute siano insoddisfacenti e che alcune scelte degli iperparametri delle prior di quest'ultimo modello siano poco ragionevoli. In particolare, riteniamo che gli intervalli di credibilità per determinati quantili della distribuzione del tempo di vita siano troppo larghi, e che alcune prior siano eccessivamente informative. Quindi abbiamo deciso di formulare dei modelli più flessibili per ottenere delle stime dei quantili di interesse più robuste e più accurate. Due sono i modelli bayesiani semiparametrici considerati. Nel primo l'errore (di regressione) è rappresentato da una mistura di distribuzioni parametriche in cui la misturante è un processo NGG; la distribuzione degli effetti casuali è parametrica, come in Leon et al. [24]. Nel secondo modello l'errore è stato trattato parametricamente come in Leon et al. [24], mentre gli effetti casuali sono stati considerati in modo non parametrico, utilizzando sempre una prior di tipo NGG. Infine abbiamo validato i risultati ottenuti utilizzando la funzione *DPglm* nel pacchetto *DPpackage* del software *R*. Si tratta di un pacchetto che permette di fare inferenza bayesiana su modelli semi-parametrici; in particolare con la funzione *DPglm* si può fare inferenza per modelli GLMM in cui gli effetti casuali hanno una distribuzione iniziale di tipo processo di Dirichlet. Abbiamo utilizzato errori di tipo Gamma, dato che la Weibull non appartiene alla famiglia esponenziale e non è contemplata in questo pacchetto; ma sottolineiamo che abbiamo usato questo package solo per fare un'analisi comparativa.

I modelli che abbiamo qui sviluppato non solo sono nuovi perchè mai applicati a questo specifico data-set, ma risultano originali anche nel contesto più generale di affidabilità bayesiana non parametrica. Infatti, a differenza dell'ambito biostatistico, la letteratura bayesiana nonparametrica in affidabilità non è molto estesa. Inoltre nella parte applicativa della tesi, abbiamo autonomamente calcolato analiticamente le espressioni delle full-conditional necessarie per poter costruire un algoritmo MCMC che permetta di campionare dalla distribuzione a posteriori; successivamente abbiamo implementato gli algoritmi nel linguaggio *C* ed eseguito molte simulazioni numeriche. Per quanto riguarda il primo modello, ad ogni iterazione dell'algoritmo simuliamo traiettorie dal processo NGG.

Gli algoritmi che simulano intere traiettorie di misure di probabilità aleatorie sono stati proposti recentemente in letteratura; sono in generale molto onerosi dal punto di vista computazionale e di difficile implementazione, ma forniscono un'informazione decisamente più ricca sul modello. Nello specifico per simulare le traiettorie del processo NGG è stato necessario invertire la funzione gamma troncata, che è nota per le sue instabilità numeriche; per fare ciò in maniera efficace abbiamo utilizzato la libreria di  $C$  chiamata *pari*. Nel secondo modello abbiamo utilizzato un algoritmo di tipo *Polya-Urn* che marginalizza rispetto al processo NGG, e che quindi risulta computazionalmente meno pesante.

Nel Capitolo 1, dopo una breve introduzione all'approccio bayesiano non parametrico, vengono presentate le principali proprietà del processo di Dirichlet. Segue la costruzione del processo NGG a partire da misure di probabilità aleatorie a incrementi indipendenti. In ultimo vengono descritte le proprietà del processo NGG.

Nel Capitolo 2 vengono introdotti i GLMM. Questa classe di modelli nasce dall'unione dei modelli lineari ad effetti misti in cui la risposta ha distribuzione gaussiana e gli effetti sono sia fissi che casuali, e i modelli lineari generalizzati che hanno risposta appartenente alla classe della famiglia esponenziale ed effetti fissi. Ci concentriamo in particolare sui modelli lineari generalizzati a effetti misti in ambito bayesiano: partiamo dal caso parametrico proposto da Zeger e Karim [38]; per poi giungere a quello semi-parametrico di Kleinmann e Ibrahim [23], in cui gli effetti casuali hanno prior nonparametrica di tipo processo di Dirichlet.

Nel Capitolo 3 viene introdotta qualche nozione relativa all'analisi di sopravvivenza (ricordiamo che in ambito ingegneristico questa prende il nome di analisi di affidabilità), e poi viene presentato il data-set relativo ai recipienti in pressione dello Space-Shuttle. In ultimo vengono introdotti i modelli AFT bayesiani menzionati in questa Introduzione: il modello parametrico di Leon et al. [24], il modello con errori nonparametrici, il modello con effetti casuali non parametrici e quello con errori gamma del pacchetto *DPpackage*.

Nel Capitolo 4 presentiamo e confrontiamo i risultati ottenuti dai vari modelli. Per ogni modello discutiamo la scelta degli iperparametri della prior, forniamo gli intervalli di credibilità dei quantili di interesse della distribuzione del tempo di vita per un nuovo recipiente in pressione, le stime a posteriori dei parametri del modello e l'analisi dei residui.

Le Appendici A e B riportano le espressioni analitiche delle full-conditional rispettivamente del modello con errori non parametrici e del modello con effetti casuali non parametrici.

# Introduction

In this work we provide a methodological study about Bayesian nonparametric random-effects models, and an application of these models in reliability.

Random-effects models separate the variability among different statistical units (or groups) from the variability inside each statistical unit. These models are also called hierarchical models since they define two or more levels of variability. For instance, they are often used in clinical trials where several observations (measurements) for each patient are taken to estimate differences among patients and the variability within each patient; in this case the grouping factor is the patient himself. These modeling techniques have application in several fields, reliability analysis among others. A wide class of random-effects models is that of the generalized linear mixed-effects models (GLMMs), where the effects are both fixed and random, and the outcome belongs to the exponential family. There are several, and often conflicting, definitions of fixed and random-effects. By Gelman [13], fixed-effects are identical for all groups in a population, while random-effects are allowed to differ from group to group.

The Bayesian approach to random effects have several advantages with respect to the frequentist one. First, frequentist inference is often based on asymptotic assumptions, while in the Bayesian framework it is always possible to make exact inference using Markov Chain Monte Carlo (MCMC) numerical integration methods, for data-set of any dimension. Moreover, in classical statistics the random-effects parameters are generally considered mutually independent; on the other hand in Bayesian statistics they are assumed exchangeable and this dependence among the effects brings to more accurate estimates. In particular, exchangeability enables to borrow strength across random-effects and so we obtain reasonable estimates also for the statistical units with few observations. We recall that the exchangeability assumption is reasonable here, since the statistical units, represented by the random-effects, could be considered as a sample without replacement from the population of statistical units.

Often it is not reasonable to make any parametric assumption about the random-effects' distribution and so, in these cases, a parametric model could bring to biased estimates of the parameters. The nonparametric approach relaxes the parametric hypothesis



for the random-effects and leads to a more robust inference.

The drawback of the Bayesian nonparametric approach consists in its computational heaviness, but the increasing computational power of the last years has made Bayesian nonparametric inference feasible and more popular in literature. However, the computational efforts required for the inference is still one limitation of Bayesian nonparametrics and hence the research of efficient algorithms is always up-to-date. See Müller and Quintana [29] for a presentation of the most common classes of nonparametric priors and of the main Bayesian nonparametric inference's techniques. Recently, Bayesian nonparametrics has become particularly popular in biostatistics; see Dunson [7] for an exhaustive description of the latest Bayesian nonparametric models in this field. In this work we focus on the Normalized Generalized Gamma (NGG) process prior. This process was introduced by Brix [3] in its unnormalized version, while Regazzini et al. [35] studied it as a prior for the first time. The NGG process can be constructed by normalization of a completely random measure, and the Dirichlet process is recovered as a particular case. Like the Dirichlet process, the NGG process selects discrete distributions with probability one, and induces a random partition on the positive integers; in particular, this means that, given a sample of size  $n$  from this process, among the  $n$  observations there could be repeated values, and these repeated values define a partition of  $\{1, \dots, n\}$ . This random partition is ruled by two parameters in the NGG case, while the grouping of the Dirichlet process is ruled by only one parameter. This additional degree of freedom makes the clustering of the NGG process prior more flexible.

The accelerated life tests (ALTs) are very important in reliability. ALT testing involves acceleration of failure times with the purpose of predicting the life-time of the product at normal use conditions. Data from an ALT can be analyzed by a so-called Accelerated Failure Time (AFT) model, where the dependence between the logarithm of the failure time is related to some explanatory variables like pressure and temperature, among others. Notice that in log-scale this is a regression model. A common choice in AFT models is the Weibull distribution for the life-time, because it has a natural interpretation of the shape and scale parameters. The AFT model is not exactly a GLMM, since the Weibull distribution does not belong to the exponential family, but it can be considered a straightforward generalization. Here, we will analyze an AFT made by NASA on some pressure vessels, which are critical components of the Space-Shuttle. The data-set consists of 108 lifetimes of pressure vessels wrapped with Kevlar yarn. The fiber of Kevlar comes from 8 different spools and 4 levels of pressure stress are used. The reliability of the pressure vessels is still relevant: in 2009 NASA organized the *Composite Pressure Vessel and Structure Summit* [5] to discuss the state of the art on this subject.

In this work, we model the failure times via semi-parametric Bayesian AFT models, and provide posterior estimates of the regression parameters and credibility intervals of

some given quantiles of the failure-time distribution for pressure vessel wrapped with fiber from a new random spool. We consider the spool effect as random, since several observations for each spool are provided, while the pressure stress level is treated as a fixed effect. Several authors studied this data-set with a frequentist approach, where both spool and pressure are considered as fixed effects; more recently Leon et al. [24] fitted a Bayesian parametric AFT model with mixed-effects. Their estimates are not completely reliable and some of their prior assumptions are questionable. According to our opinion, some of their interval estimates of given quantiles of the failure-time distribution are too large, the hyperparameters of the prior are too much informative, and their posterior estimates are affected by a large Monte Carlo error. Hence, we assumed more flexible models with the purpose of getting more accurate and more robust estimates. We studied two different Bayesian semiparametric models. In the first model, the regression error is represented by a mixture of parametric distribution with a NGG mixing measure, while the distribution of the random-effects is parametric as in Leon et al. [24]. In the second model, the error is treated parametrically as in Leon et al. [24], while we consider nonparametric random-effects, using a NGG process prior. Finally, we validated our results using the function *DPglm* of the package *DPpackage* in the statistical software *R*. This package offers several functions to fit Bayesian nonparametric and semiparametric models; in particular the function *DPglm* fits GLMMs, where the random-effects have Dirichlet process prior distribution. The outcome is assumed here Gamma-distributed, since the Weibull distribution does not belong to the exponential family and it is not available in this package. We underline that we used *DPpackage* only for comparative purposes only.

Not only these models are new for this given data-set, but also in the more general framework of reliability under a Bayesian approach. In fact, despite biostatistics, reliability nonparametric Bayesian literature is not particularly wide. Moreover, we derived the analytical expressions of the full-conditional distributions needed to make a MCMC algorithm to sample from the posterior distribution; then we coded the algorithms in the programming language *C* and we made several numerical simulations. At each iteration of the first model's algorithm, we sample a trajectory of the NGG process. The algorithms that simulate complete trajectories of random probability measures are quite recent in literature; they are particularly time consuming and hard to implement, but they provide much information on the model. In particular, to sample a trajectory of the NGG process, we must invert the truncated gamma function, which is well-know for its numerical instability (we used a *C* library called *pari*). In the second model, we implemented a *Polya-urn scheme* algorithm, which integrates out the NGG process, and so it is computationally less heavy.

In Chapter 1, after a brief introduction to nonparametric Bayesian approach, we present the main properties of the Dirichlet process. Then a construction of the NGG

process by normalization of a completely random measure is provided. Finally, we introduce the main properties of the NGG process.

In Chapter 2 we present the GLMM. This class of models can be seen as the generalization of mixed-effects linear models and generalized linear models. We will focus on the Bayesian approach to GLMM: first, we introduce the parametric model of Zeger and Karim [38]; then the semi-parametric one of Kleinmann and Ibrahim [23], where the random effects have Dirichlet process prior.

In Chapter 3 we provide basic notions of survival analysis (or reliability analysis in engineering), and then we introduce the data-set of the NASA pressure vessels. Finally, we present the Bayesian AFT models mentioned before: Leon et al.[24]'s parametric model, the model with nonparametric error, the model with nonparametric random-effects, and the model with Gamma-distributed error of *DPpackage*.

In Chapter 4 we present and compare the results of the different models. For each of them, we discuss the prior hyperparameter's choice, and we provide the interval estimates of given quantiles of the failure time distribution for a new pressure vessel wrapped with fiber from a new random spool, the posterior estimates of the parameters and the analysis of the residuals.

In Appendix A and B we provide the analytic expressions of the full-conditionals of the model with nonparametric error and those of the model with nonparametric random-effects, respectively.

# Chapter 1

## Bayesian Nonparametrics

### 1.1 Exchangeability assumption

Classical statistics is based on a framework where observations  $X_1, X_2 \dots$  are assumed independent and identical distributed (i.i.d.) from a unknown probability distribution  $P$ . We say that we are considering a parametric framework when  $P$  belongs to a parametric family, otherwise we are considering a nonparametric framework when  $P$  lies in the space of probability distributions  $\mathcal{P}(\mathbb{R})$ .

It is possible to distinguish the two cases also in the Bayesian setting. In the parametric case we have a prior  $\Pi$  on a finite dimensional space  $\Theta$  and, given  $\theta$ , the observations are assumed i.i.d. from  $P_\theta$ . In the nonparametric case, we have a prior  $\Pi$  on the space  $\mathcal{P}(\mathbb{R})$  of all probability distributions on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and, given  $P$ , the observations are assumed i.i.d. from  $P$ .

Under the assumption of exchangeability, de Finetti's Representation Theorem gives a validation of the Bayesian setting.

Let consider an infinite sequence of observations  $(X_n)_{n \geq 1}$  defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , with each  $X_i$  taking values on  $\mathbb{R}$  endowed with the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$ . This last hypothesis can be relaxed and we could consider observations which take values in a complete metric and separable space  $\mathbb{X}$ . In this work it is enough to consider  $\mathbb{X} = \mathbb{R}$ .

**Definition 1.** *A sequence  $(X_n)_{n \geq 1}$  is exchangeable when, for any finite permutation  $\pi$  of  $(1, 2, \dots, n)$ , the random vectors  $(X_1, \dots, X_n)$  and  $(X_{\pi(1)}, \dots, X_{\pi(n)})$  have the same probability distribution.*

There are several types of dependence among a sequence of observations  $(X_n)_{n \geq 1}$ . Under the exchangeability assumption, the information that the observations  $X_i$ s provide is independent of the order in which they are collected. For instance, if we sample without

replacement from an urn with infinite marbles of different colors, the sequence of colors that we obtain is exchangeable.

A random element defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ , with values in  $\mathcal{P}(\mathbb{R})$ , is called *random probability measure* (r.p.m.).

With a view to our utilization of r.p.m.s in statistics, the two main desirable properties for the class of r.p.m. are a large support, and a posterior distribution that is analytically tractable. A prior with a large support is an obvious requirement, and a tractable posterior reduces the computational complexity. In fact computational heaviness is still one limitation of Bayesian nonparametrics.

The most popular r.p.m.s in literature are Dirichlet Processes, Polya Trees and Bernstein Polynomials. A recent review of the main r.p.m. classes appears in Müller and Quintana [29].

Now we give formal definitions of the Borel  $\sigma$ -algebra on  $\mathcal{P}(\mathbb{R})$  introducing the topology of weak convergence. The space  $\mathcal{P}(\mathbb{R})$  is equipped with the topology of the weak convergence which makes it a complete and separable metric space. We will write that  $P_n \xrightarrow{w} P$  ( $P_n$  converges weakly to  $P$ ), if

$$\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP, \text{ as } n \rightarrow +\infty$$

for all bounded continuous function  $f$  on  $\mathbb{R}$ . For any  $P_0$  a neighborhood base consists of sets of the form

$$\cap_{i=1}^k \{P : |\int f_i dP_0 - \int f_i dP| < \epsilon\}$$

where  $f_i, i = 1, \dots, k$  are bounded continuous function on  $\mathbb{R}$ ,  $k \geq 1$  and  $\epsilon > 0$ .

The Borel  $\sigma$ -algebra on  $\mathcal{P}(\mathbb{R})$  is the smallest  $\sigma$ -algebra generated by the open sets in the weak topology.

**Theorem 1** (de Finetti). *The sequence  $(X_n)_{n \geq 1}$  is exchangeable if, and only if, there exists a unique probability measure  $q$  on  $\mathcal{P}(\mathbb{R})$  such that, for any  $n \geq 1$  and any Borel sets  $B_1, B_2, \dots, B_n$ ,*

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \int_{\mathcal{P}(\mathbb{R})} \prod_{i=1}^n p(B_i) q(dp).$$

*Equivalently,*

$$\begin{aligned} X_1, \dots, X_n | P &\stackrel{\text{iid}}{\sim} P \\ P &\sim q(\cdot). \end{aligned}$$

In the parametric case  $q$  is concentrated on a parametric family

$$\begin{aligned} X_1, \dots, X_n | \theta &\stackrel{\text{iid}}{\sim} f_\theta(\cdot) \\ \theta &\sim \pi(\cdot), \end{aligned}$$

where  $X_i | \theta$  and  $\theta$  are absolutely continuous (with respect to the Lebesgue measure) or discrete probability distributions, for  $i = 1, \dots, n$ .  $f_\theta(\cdot)$  and  $\pi(\cdot)$  are the probability density functions of  $X_i | \theta$  and  $\theta$  respectively.

By Bayes' Theorem, the posterior probability distribution of  $\theta$ , i.e. the conditional probability distribution of  $\theta$  given  $X_1, \dots, X_n$ , has probability density function

$$\pi(\theta | X_1 = x_1, \dots, X_n = x_n) = \frac{\prod_{i=1}^n f_\theta(x_i) \pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f_\theta(x_i) \pi(d\theta)}.$$

The predictive distribution of a new observation  $X_{n+1}$  has probability density function

$$f_\theta(x | X_1 = x_1, \dots, X_n = x_n) = \int_{\Theta} f_\theta(x) \pi(d\theta | X_1 = x_1, \dots, X_n = x_n)$$

Similarly we can make inference and prediction in the nonparametric setting. In this case  $q$  is a probability measure on  $\mathcal{P}(\mathbb{R})$ , the posterior distribution can be derived by

$$\mathcal{L}(dP | X_1, \dots, X_n) = \frac{\prod_{i=1}^n \mathcal{L}(X_i | P) \mathcal{L}(dP)}{\int_{\mathcal{P}(\mathbb{R})} \prod_{i=1}^n \mathcal{L}(X_i | P) \mathcal{L}(dP)}$$

and the predictive distribution of a new observation  $X_{n+1}$  is

$$\mathcal{L}(X_{n+1} | X_1, \dots, X_n) = \int_{\mathcal{P}(\mathbb{R})} \mathcal{L}(X_{n+1} | P) \mathcal{L}(dP | X_1, \dots, X_n).$$

## 1.2 Dirichlet processes

The Dirichlet process is a useful family of prior distributions on  $\mathcal{P}(\mathbb{R})$  introduced by Ferguson [11]. The Dirichlet prior is easy to elicit, has a manageable posterior and other nice properties. It can be viewed as an infinite-dimensional generalization of the finite-dimensional Dirichlet distribution.

**Definition 2.** Let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  with  $\alpha_i > 0$  for  $i = 1, 2, \dots, k$ . The random vector  $P = (P_1, P_2, \dots, P_k)$ ,  $\sum_{i=1}^k P_i = 1$ , has Dirichlet distribution with parameter  $\alpha$ , if  $(P_1, \dots, P_{k-1})$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^{k-1}$  with density

$$f(p_1, p_2, \dots, p_{k-1}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\Gamma(\alpha_1) \Gamma(\alpha_2) \dots \Gamma(\alpha_k)} p_1^{\alpha_1-1} p_2^{\alpha_2-1} \dots p_{k-1}^{\alpha_{k-1}-1} (1 - \sum_{i=1}^{k-1} p_i)^{\alpha_k-1},$$

where  $0 \leq p_i \leq 1 \forall i$ ,  $0 \leq p_1 + \dots + p_{k-1} \leq 1$ , 0 otherwise.

We will write  $P \sim D(\alpha)$ .

**Definition 3.** Let  $\alpha$  a finite measure on  $\mathbb{R}$ ,  $a =: \alpha(\mathbb{R})$ ; let  $\alpha_0(\cdot) = \alpha(\cdot)/a$ . A r.p.m.  $P$  with values in  $\mathcal{P}(\mathbb{R})$  is a Dirichlet process on  $\mathbb{R}$  with parameter  $\alpha$  if, for any finite measurable partition  $B_1, \dots, B_k$  of  $\mathbb{R}$ ,

$$(P(B_1), \dots, P(B_k)) \sim D(\alpha(B_1), \dots, \alpha(B_k)).$$

We will write  $P \sim DP(\alpha)$  for short. It can be proved that such a process exists (see Ferguson [11]). If  $P \sim DP(\alpha)$ , it follows that  $\mathbb{E}[P(A)] = \alpha_0(A)$  for any Borel set  $A$ , and thus we can say that  $\alpha_0$  is the prior expectation of  $P$ .

The Dirichlet prior is a conjugate prior on  $\mathcal{P}(\mathbb{R})$ ; in fact, let  $(X_1, X_2, \dots, X_n)$  be a sample from a Dirichlet process  $P$ , i.e.

$$\begin{aligned} X_1, X_2, \dots, X_n | P &\stackrel{\text{iid}}{\sim} P \\ P &\sim DP(\alpha). \end{aligned}$$

Then the posterior distribution of  $P$ , given  $X_1, X_2, \dots, X_n$ , is

$$P | X_1, X_2, \dots, X_n \sim DP(\alpha + \sum_{i=1}^n \delta_{X_i}).$$

In this case, it can be proved that the distribution of  $X_{n+1}$  can be described as follows:

$$X_1 \sim \alpha_0 \tag{1.1}$$

$$X_{n+1} | X_1, \dots, X_n \sim \frac{a}{a+n} \alpha_0 + \frac{n}{a+n} \left( \frac{\sum_{i=1}^n \delta_{X_i}}{n} \right) \tag{1.2}$$

Notice that the predictive distribution in (1.1), called *Blackwell-MacQueen Urn Scheme*, is a mixture of the base-line measure  $\alpha_0$  and the previous observations. This means that there is a positive probability of coincident values for any finite and positive  $a$ . Moreover if  $\alpha_0$  is an absolutely continuous probability measure, then  $X_{n+1}$  will assume a different, distinct value with probability  $\frac{a}{a+n}$ . Formula (1.1) allows us to sample (marginally) from  $P$  without simulating any trajectory of the Dirichlet process.

Let  $(X_1, X_2, \dots, X_n)$  be a sample from  $P$ , where  $P \sim DP(\alpha)$ . If  $K_n$  denotes the random variable representing the number of distinct values among  $(X_1, X_2, \dots, X_n)$ . Antoniak (1974) proved that the distribution of  $K_n$  is the following

$$\mathbb{P}(K_n = k) = c_n(k) n! a^k \frac{\Gamma(a)}{\Gamma(a+n)}, \quad k = 1, 2, \dots, n, \tag{1.3}$$

where  $c_n(k)$  is the absolute value of Stirling number of the first kind, which can be tabulated or computed by a software. From (1.3) it is clear that the mass parameter  $a$  influences the prior on the number of clusters. Larger  $a$  gives rise to a higher prior number of components. .

Sethuraman (1994) provided a useful representation of the Dirichlet process. Its construction gives an insight on the structure of the process and provides an easy way to simulate its trajectories.

Let consider two independent sequences of random variables  $(\theta_i)_{i \geq 1}$  and  $(\tau_i)_{i \geq 1}$  such that  $\theta_i \stackrel{\text{iid}}{\sim} \text{beta}(1, a)$  and  $\tau_i \stackrel{\text{iid}}{\sim} \alpha_0$  (defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ), and define the following weights

$$\begin{cases} p_1 = \theta_1 \\ p_n = \theta_n \prod_{i=1}^{n-1} (1 - \theta_i), \quad n \geq 2 \end{cases}$$

It is straightforward to see that  $0 \leq p_n \leq 1$   $n = 1, 2, \dots$  and  $\sum_{n=1}^{\infty} p_n = 1$  a.s..

This construction is called *stick-breaking*. In fact  $p_1$  represents a piece of a unit-length stick,  $p_2$  represents a piece of the remainder of the stick and so on, where each piece is independently modeled as a  $\text{beta}(1, a)$  random variable scaled down to the length of the remainder of the stick.

Now we can define a random variable  $P$  on  $\mathcal{P}(\mathbb{R})$

$$P(A) = \sum_{n=1}^{\infty} p_n \delta_{\tau_n}(A), \quad A \in \mathcal{B}(\mathbb{R}).$$

Sethuraman (1994) proved that  $P$  has Dirichlet prior distribution, i.e.  $P$  is a Dirichlet process with parameter  $\alpha$ . From this construction it is clear that a Dirichlet process has discrete trajectories, i.e. if  $P \sim DP(\alpha)$ ,  $\mathbb{P}(\{\omega : P(\omega) \text{ is discrete}\}) = 1$ .

As mentioned in Section 1.1,  $\mathcal{P}(\mathbb{R})$  is a complete separable metric space, and hence any probability measure  $\Pi$  on  $\mathcal{P}(\mathbb{R})$  has a support (the smallest closed set of measure 1). Let  $E$  be the support of the finite measure  $\alpha$  on  $\mathbb{R}$ . Then it can be shown that  $M_\alpha = \{P : \text{support of } P \subset E\}$  is the weak support of  $DP(\alpha)$ , i.e. the set of all the probability distributions with support contained in the support of the measure  $\alpha$  is the weak support of  $DP(\alpha)$ .

### 1.3 Normalized random measures with independent increments

The class of the *normalized random measures with independent increments* (NRMIs) generalizes the Dirichlet process. The NRMIs are defined as normalization of *completely random measures*, that can be constructed from Poisson processes. A nonparametric Bayesian analysis of NRMIs is developed in James et al. [18] and for a comprehensive introduction to Poisson processes and completely random measures we refer to Kingman [22].

Before constructing the NRMIs we introduce some preliminary concepts. First we give the definition of a Poisson process on a general space  $\mathbb{S}$ , then we provide the definition



of a completely random measure and finally we show how to construct a completely random measure from a specific Poisson process.

**Definition 4.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be some probability space,  $\mathbb{S}$  be complete separable metric space, and  $\nu$  a non-atomic measure on  $\mathbb{S}$ . A Poisson Process  $N$ , with state space  $\mathbb{S}$  and intensity measure  $\nu$ , defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ , is a stochastic process such that:

1. for any disjoint measurable subset  $A_1, A_2, \dots, A_n$  of  $\mathbb{S}$ , the random variables  $N(A_1), N(A_2), \dots, N(A_n)$  are mutually independent;
2. for any  $A$  in  $\mathbb{S}$ ,  $N(A) \sim \text{Poisson}(\nu(A))$ .

A completely random measure  $\Phi$  on  $\mathbb{R}$  is a random measure such that, for any collection of disjoint measurable subsets  $A_1, A_2, \dots$  of  $\mathbb{R}$ , the random variables  $\Phi(A_i)$  are independent. Hence a Poisson process is a completely random measure with Poisson's finite dimensional distribution.

Now let us consider  $\mathbb{S} = \mathbb{R}^+ \times \mathbb{R}$  and take a measure  $\nu$  such that  $\int_0^\infty \min(s, 1)\nu(ds, \mathbb{R}) < +\infty$ . We can construct a completely random measure  $\Phi$  as a linear functional of the Poisson random measure  $N$ , with space state  $\mathbb{S}$  and intensity  $\nu$ , as

$$\Phi(B) = \int_{\mathbb{R}^+ \times B} sN(ds, dx), \quad B \in \mathcal{B}(\mathbb{R}). \quad (1.4)$$

It can be proved that  $\Phi$  is a completely random measure on  $\mathbb{R}$  (see Kingman [22]). Moreover  $\Phi$  is a purely atomic measure and in fact its atoms correspond to the points of  $N$ : if  $(s, x)$  is a point of  $N$ , then  $\Phi$  has an atom of weight  $s$  at  $x$ .

By Campbell's theorem, the moment generating function of  $\Phi(B)$  for any measurable  $B$  is

$$\mathbb{E}[e^{-t\Phi(B)}] = \exp\left\{\int_0^\infty (e^{-ts} - 1)\nu(ds, B)\right\}, \quad t > 0.$$

Any completely random measure  $\Phi$  so defined is identified by its corresponding intensity measure  $\nu$ . A random probability measure is called *homogeneous* if its intensity measure  $\nu$  can be decomposed in  $\nu(ds, dx) = \rho(ds)\gamma(dx)$ , where  $\rho$  and  $\gamma$  are measures on  $\mathbb{R}^+$  and  $\mathbb{R}$ , respectively.

In formula (1.4) we imposed that  $\int_0^\infty \min(s, 1)\nu(ds, \mathbb{R}) < +\infty$ ; adding the condition that  $\nu(\mathbb{S}) = +\infty$ , Regazzini et al. [35] showed that  $\mathbb{P}(0 < T := \Phi(\mathbb{R}) < +\infty) = 1$ . Therefore we define a r.p.m. process  $P$  as

$$P(B) = \frac{\Phi(B)}{T}, \quad B \in \mathcal{B}(\mathbb{R}).$$

$P$  is called *normalized random measure with independent increments* (NRMI) on  $\mathbb{R}$ . By the discreteness of  $\Phi$ , it follows that the NRMI  $P$  selects discrete distributions almost

surely. Moreover it admits a series representation of the kind

$$P = \sum_{i=1}^{\infty} p_i \delta_{\tau_i}.$$

If the underlying intensity measure is homogeneous,  $p_i$  and  $\tau_i$  reciprocally independent.

The Dirichlet process  $DP(\alpha)$  can be obtained by normalization of a gamma random measure, where the underlying intensity measure is  $\nu(A, B) = \alpha(B) \int_A s^{-1} e^{-s} ds$ ,  $A \in \mathcal{B}(\mathbb{R}^+)$  and  $B \in \mathcal{B}(\mathbb{R})$ .

Other examples of NRMI are normalized inverse-gaussian processes and normalized generalized gamma processes. We will present the latter in the Subsection 1.4 with more details.

## 1.4 Normalized generalized gamma processes

The *normalized generalized gamma process* (NGG) is a random probability measure with values in  $\mathcal{P}(\mathbb{R})$  constructed via normalization of a generalized gamma random measure.

As presented in Subsection 1.2, the clustering behavior of the Dirichlet process is controlled by the parameter  $a$ . The NGG process has an additional parameter  $\sigma$  belonging to  $[0, 1]$  that reinforces the clustering mechanism, and includes the Dirichlet process as a particular case ( $\sigma = 0$ ).

The generalized gamma random measure has been introduced by Brix [3], while an introduction to the NGG process can be found in Lijoi and Prünster [27], see Lijoi et al. [26] and Argiento et al. [1] for applications of NGG processes in Bayesian nonparametric mixture models.

Let consider a Poisson process  $N$  with space state  $\mathbb{S} = \mathbb{R}^+ \times \mathbb{R}$  and mean measure

$$\nu(A, B) = \kappa(B) \int_A \rho(ds), \quad A \in \mathcal{B}(\mathbb{R}^+), \quad B \in \mathcal{B}(\mathbb{R}),$$

where  $\kappa(\cdot)$  is an absolutely continuous finite measure on  $\mathbb{R}$  and

$$\rho(ds) = \frac{1}{\Gamma(1-\sigma)} s^{-\sigma-1} e^{-\omega s} ds, \quad s > 0, \quad \text{where } \omega \geq 0, 0 < \sigma \leq 1. \quad (1.5)$$

With this choice of  $\nu$  and  $\sigma$ , define the completely random measure  $\Phi$  as in (1.4). For any measurable  $B$ , the moment generating function of  $\Phi(B)$  is

$$\mathbb{E}[e^{-t\Phi(B)}] = \exp\left\{-\frac{\kappa(B)}{\sigma} [(\omega + t)^\sigma - \omega^\sigma]\right\}, \quad t > 0.$$

Since  $\int_0^\infty \min(s, 1) \nu(ds, \mathbb{R}) < +\infty$  and  $\nu(\mathbb{S}) = +\infty$ , we define a *normalized generalized gamma process*  $P$  via normalization of the completely random measure  $\Phi$  defined above.

Hence a NGG process  $P$  is characterized by the set of parameters  $(\sigma, \kappa(\cdot), \omega)$ , when  $0 < \sigma \leq 1$ . It can be showed that this parameterization is not unique, since  $(\sigma, \kappa(\cdot), \omega)$  and  $(\sigma, s^\sigma \kappa(\cdot), \omega/s)$  yield the same distribution for  $P$ , for any  $s > 0$  (see Pitman [32]). For short we will write  $P \sim NGG(\sigma, \omega, \kappa, P_0)$ , where  $\kappa := \kappa(\mathbb{R}^+)$  and  $P_0 = \kappa(\cdot)/\kappa(\mathbb{R}^+)$ ,  $\omega > 0$ , and  $\kappa$  is an absolutely continuous (with respect to the Lebesgue measure) finite measure on  $\mathbb{R}$ .

As mentioned in Subsection 1.3, the NGG process  $P$  admits a series representation:

$$P = \sum_{i=1}^{\infty} p_i \delta_{\tau_i} = \sum_{i=1}^{\infty} \frac{J_i}{T} \delta_{\tau_i}, \quad (1.6)$$

where the sequences  $(p_i)_{i \geq 1}$  and  $(\tau_i)_{i \geq 1}$  are independent. Moreover  $p_i := \frac{J_i}{T}$ , where  $(J_i)_{i \geq 1}$  are the ranked points of a Poisson process on  $\mathbb{R}^+$  with mean intensity  $\rho(ds)$ ,  $T = \sum_{i=1}^{\infty} J_i$ , and  $\tau_i$  are i.i.d. from  $P_0$ . Since the  $J_i$  are ranked points,  $P_1 \geq P_2 \geq \dots$ , and we notice that  $\mathbb{P}(\sum_{i=1}^n p_i = 1) = 1$  by definition.

Unlike the Dirichlet process, the NGG does not admit any analytic expression of its finite dimensional distributions. Nonetheless the mean, the variance of  $P(B)$ , for any measurable  $B$ , are

$$\begin{aligned} \mathbb{E}[P(B)] &= P_0(B), \\ \text{Var}[P(B)] &= P_0(B)(1 - P_0(B))\mathcal{I}(\sigma, \kappa), \end{aligned} \quad (1.7)$$

and, for any measurable  $B_1$  and  $B_2$ ,

$$\text{Cov}(P(B_1), P(B_2)) = (P_0(B_1 \cap B_2) - P_0(B_1)P_0(B_2))\mathcal{I}(\sigma, \kappa),$$

where

$$\mathcal{I}(\sigma, \kappa) := \left(\frac{1}{\sigma} - 1\right) \left(\frac{\kappa}{\sigma}\right)^{1/\sigma} e^{\frac{\kappa}{\sigma}} \Gamma\left(\frac{1}{\sigma}, \frac{\kappa}{\sigma}\right) = \left(\frac{1}{\sigma} - 1\right) \int_1^{\infty} e^{\frac{\kappa}{\sigma}(y-1)} y^{-\frac{1}{\sigma}-1} dy$$

and  $\Gamma(\alpha, x) := \int_x^{\infty} e^{-t} t^{\alpha-1} dt$  is the incomplete gamma function. It can be proved that  $P(B) \xrightarrow{w} P_0(B)$  if  $\sigma \rightarrow 1$  or  $\kappa \rightarrow +\infty$ ; and that  $P(B) \xrightarrow{w} \delta_{\tau}(B)$  if  $\sigma \rightarrow 0$  and  $\kappa \rightarrow 0$ , where  $\tau \sim P_0$ .

If we fix  $\sigma = 0$  and  $\kappa > 0$ , the underlying intensity measure is  $\nu(A, B) = \kappa(B) \int_A s^{-1} e^{-s} ds$  for any measurable  $A$  and  $B$ , and we recover the Dirichlet process  $DP(\kappa P_0)$ .

Since the NGG selects discrete distributions a.s., sampling from  $P$  induces a random partitions  $\Pi$  on the positive integers and thus can be considered as a particular case of the so-called *species sampling model*. This class of r.p.m. was introduced by Pitman [31].

We recall that a partition of a set  $\mathbb{X}$  is a set of nonempty subsets of  $\mathbb{X}$  such that each element  $X \in \mathbb{X}$  is in exactly one of these subsets. Therefore a random partition of a set  $\mathbb{X}$  is a random variable with values in the space of the partitions of  $\mathbb{X}$ . Let  $X_1, X_2, \dots$ , given  $P$ , be i.i.d. from  $P$ , and  $\mathbb{X} = \{X_1, X_2, \dots, X_n\}$ , for any  $n \geq 1$ ;  $X_i$  and  $X_j$  belong to the

same subset of the partition if, and only if,  $X_i = X_j$ , where  $i, j = 1, 2, \dots$ . Since  $(X_i)_{i \geq 1}$  is exchangeable, the probability of a partition of  $(X_1, \dots, X_n)$  depends only on  $n$  and on the cardinalities of its subsets, for any  $n \geq 1$ .

Let consider, for any  $n$ ,  $(X_1, X_2, \dots, X_n)$  a sample from a NGG process  $P$ , and define  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_k)$  the vector of the distinct values among  $\{X_1, X_2, \dots, X_n\}$ ; then the marginal prior distribution of  $(X_1, X_2, \dots, X_n)$  is identified by the joint distribution of the random partition  $\Pi_n$  of  $(1, 2, \dots, n)$ , and  $\boldsymbol{\psi}$ . Indeed,

$$\begin{aligned} \mathbb{P}(\Pi_n = \pi_n, \psi_1 = B_1, \psi_2 = B_2, \dots, \psi_k = B_k) &= \mathbb{P}(\Pi_n = \pi_n) \prod_{l=1}^k P_0(B_l) \\ &= p(e_1, e_2, \dots, e_k) \prod_{l=1}^k P_0(B_l), \end{aligned} \quad (1.8)$$

where  $e_l = \#\{X_i : X_i = \psi_l, 1 \leq i \leq n\}$  and  $\sum_{l=1}^k e_l = n$ . The symmetric and non-negative function  $p$  in (1.8) is called *exchangeable partition probability function* (EPPF) and it can be proved that has the following expression in the general frame of the species sampling models:

$$p(e_1, e_2, \dots, e_k) = V_{n,k} \prod_{l=1}^k (1 - \sigma)_{e_l - 1}, \quad k = 1, \dots, n, \quad n = 1, 2, \dots$$

where  $(a)_n = a(a+1) \dots (a+n-1)$  with the convention  $(a)_0 = 1$ . In the NGG specific case  $V_{n,k}$  is

$$V_{n,k} = \frac{\sigma^{k-1} e^{\kappa/\sigma}}{\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \left(\frac{\kappa}{\sigma}\right)^{i/\sigma} \Gamma\left(k - \frac{i}{\sigma}; \frac{\kappa}{\sigma}\right). \quad (1.9)$$

The term  $V_{n,k}$  rules the prior distribution of the number of clusters, while  $\prod_{l=1}^k (1 - \sigma)_{e_l - 1}$  is responsible for the size of the clusters. We notice that the latter depends only on  $\sigma$  and that for a large value of  $\sigma$  most of the cluster will have small size. To stress the dependence on  $\sigma$  and  $\kappa$  we will write  $p(e_1, e_2, \dots, e_k; \sigma, \kappa)$ .

It can be showed that the law of the random variable representing the number of distinct values among  $(X_1, X_2, \dots, X_n)$  has the following expression

$$\mathbb{P}(K_n = k) = \frac{V_{n,k}}{\sigma^k} \mathcal{G}(n, k, \sigma) \quad k = 1, 2, \dots, n,$$

where  $\mathcal{G}(n, k, \sigma) = \frac{1}{k!} \sum_{l=0}^k (-1)^l \binom{k}{l} (-l\sigma)_n$  is the generalized Stirling number and  $V_{n,k}$  as in (1.9).

Let  $(X_1, X_2, \dots, X_n)$  be a sample from a NGG process  $P$ , i.e.

$$\begin{aligned} X_1, X_2, \dots, X_n | P &\stackrel{\text{iid}}{\sim} P \\ P &\sim \text{NGG}(\sigma, \omega, \kappa, P_0). \end{aligned}$$

Then the predictive distribution of  $X_{n+1}$ , given  $X_1, X_2, \dots, X_n$ , can be represented as follow

$$\mathbb{P}(X_{n+1} \in B | X_1, X_2, \dots, X_n) = w_0(n, k; \sigma, \kappa) P_0(B) + w_1(n, k; \sigma, \kappa) \sum_{l=1}^k (e_l - \sigma) \delta_{\psi_l}(B), \quad (1.10)$$

where

$$w_0(n, k; \sigma, \kappa) = \frac{p(e_1, \dots, e_k, 1; \sigma, \kappa)}{p(e_1, \dots, e_k; \sigma, \kappa)} \quad (1.11)$$

$$w_1(n, k; \sigma, \kappa)(e_l - \sigma) = \frac{p(e_1, \dots, e_l + 1, \dots, e_k; \sigma, \kappa)}{p(e_1, \dots, e_k; \sigma, \kappa)}, \quad (1.12)$$

for any  $k = 1, \dots, n$  (see Pitman [31]). We notice that  $w_0(n, k; \sigma, \kappa) + w_1(n, k; \sigma, \kappa) \sum_{l=1}^k (e_l - \sigma) = 1$  and that (1.11) has a mixture structure similar to (1.1). In fact with probability  $w_0(n, k; \sigma, \kappa)$   $X_{n+1}$  will be different from the previous ones, while with probability  $w_1(n, k; \sigma, \kappa)(e_l - \sigma)$  coincides with  $\psi_l$ , for  $1 \leq l \leq k$ .

Unlike the Dirichlet prior, the NGG prior is not conjugate on  $\mathcal{P}(\mathbb{R})$ . However James et al. [18] provided a posterior characterization of the NGG in the form of a mixture representation. In fact, let consider a sample  $X_1, X_2, \dots, X_n$  from a NGG process  $P = \Phi/T$ , and define a latent variable  $U := \Gamma_n/T$ , where  $\Gamma_n \sim \Gamma(n, 1)$ ; then it is possible to describe the distribution of  $\Phi$ , given  $X_1, X_2, \dots, X_n$  and the latent variable  $U = u$ , as follows

$$\Phi | (X_1, X_2, \dots, X_n, U = u) \sim \Phi_u + \sum_{l=1}^k L_l \delta_{\psi_l},$$

where  $\Phi_u$  is a generalized gamma process with parameters  $(\sigma, \kappa, \omega + u, P_0)$ , and each  $L_l$ , conditionally on  $X_1, X_2, \dots, X_n$  and the latent variable  $U = u$ , is independent of  $\Phi_u$ , and distributed as  $gamma(e_l - \sigma, \omega + u)$ , for  $1 \leq l \leq k$ . Thus the posterior distribution of  $P$ , given  $u$ , is

$$P | (X_1, X_2, \dots, X_n, U = u) \sim \frac{1}{T_u + \sum_{l=1}^k L_l} \sum_{l=1}^{\infty} J_{l,u} \delta_{\tau_l} + \frac{1}{T_u + \sum_{l=1}^k L_l} \sum_{l=1}^k L_l \delta_{\psi_l}, \quad (1.13)$$

where  $T_u := \Phi_u(\mathbb{R})$  and  $J_{l,u}$  are the jumps of a NGG( $\sigma, \kappa, \omega + u, P_0$ ) process. We notice that (1.13) is a mixture of a NGG( $\sigma, \kappa, \omega + u, P_0$ ) process and a discrete r.p.m. with fixed points equal to the observed distinct values  $\psi_1, \psi_2, \dots, \psi_k$ .

## Chapter 2

# Generalized Linear Mixed Models

The generalized linear mixed model (GLMM) unifies two classes of regression models: the linear mixed models and the generalized linear models.

A linear mixed model contains both fixed and random effects, and can be used when the data have normal errors (the difference between fixed and random effects will be explained later in this section). This model is useful when repeated measurements are made on the same statistical units. In fact it takes into account the variability between statistical units and the variability within each unit. In a generalized linear model the outcomes can be generated from any probability distribution in the exponential family, but it is usually used for uncorrelated data. We remind that the exponential family includes a wide range of both discrete and continuous probability distributions (for instance normal, gamma, Poisson, binomial). The GLMM gets through the limitations of those two models. Indeed it admits both fixed and random effects, and can be used when the data are correlated and generated from any probability distribution in the exponential family.

While GLMM frequentist inference relies on asymptotic assumptions, Zeger and Karim [38] proposed a parametric Bayesian approach and showed that exact inference for any sample size can be obtained through a MCMC method. Kleinman and Ibrahim [23] considered a Bayesian semi-parametric approach to GLMM, where the parametric (Gaussian) assumption on the distribution of the random effects is relaxed, adapting a Dirichlet process prior. They argue that a "wrong" distribution assumption on the random effects can bias the results obtained.

Recent developments of computational schemes make nonparametric Bayesian inference feasible, even if computational complexity is still a limitation in the case of a large dataset.

## 2.1 Linear mixed models

Let consider  $N$  statistical units, with  $n_i$  repeated measurements for any unit  $i$ ,  $p$  fixed effects and  $v$  random effects. For example repeated measurements on the same patient, or several patients from the same hospital. The simplest linear mixed model for the outcome  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$  is

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \quad 1 \leq i \leq N, \quad (2.1)$$

where  $\boldsymbol{\varepsilon}_i \stackrel{\text{iid}}{\sim} N_{n_i}(\mathbf{0}, \sigma^2 I_{n_i})$ ,  $\boldsymbol{\alpha}_i \stackrel{\text{iid}}{\sim} N_v(\mathbf{0}, \Lambda)$ , and  $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)$  and  $(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)$  are independent. Moreover  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}_i$  are parameter vectors (or effects) of dimension  $p$  and  $v$ , respectively, and  $X_i$  and  $Z_i$  are the design matrices of dimension  $(n_i, p)$  and  $(n_i, v)$ , respectively.

Thus we can write that

$$\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\alpha}_i \stackrel{\text{iid}}{\sim} N_{n_i}(X_i\boldsymbol{\beta} + Z_i\boldsymbol{\alpha}_i, \sigma^2 I_{n_i})$$

and

$$\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2, \Lambda \sim N_{n_i}(X_i\boldsymbol{\beta}, Z_i\Lambda Z_i' + \sigma^2 I_{n_i})$$

The coefficient  $\boldsymbol{\beta}$  represents the population-mean effects, while  $\boldsymbol{\alpha}_i$  represents the effect of unit  $i$ . We notice also that each unit  $i$  has its own covariance structure  $Z_i\Lambda^{-1}Z_i' + \sigma^2 I_{n_i}$ . Hence a mixed-effects model is nothing else but a multilevel model, where there are the units at the first level, and the repeated measurements at a second one.

## 2.2 Generalized linear mixed models

Before adding random effects into a generalized linear model, we provide some details about the exponential family.

A probability distribution belongs to the univariate exponential family if its density can be written as follow

$$f(y|\theta, \tau) = \exp\{[y\theta - a(\theta)]/\tau + c(y, \tau)\},$$

where

$$\mathbb{E}[Y|\theta, \tau] = \frac{da(\theta)}{d\theta}$$

and

$$\text{Var}[Y|\theta, \tau] = \tau \frac{d^2a(\theta)}{d\theta^2},$$

where  $\theta \in \mathbb{R}$  is called canonical parameter and  $\tau > 0$  is the dispersion parameter. It is possible to provide a multivariate definition, but in this work it is enough to consider the one-dimensional case.

**Example 1.** A random variable  $Y \sim \text{gamma}(\alpha, \beta)$  (gamma-distributed with shape  $\alpha$  and rate  $\beta$ ) has density function

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad y \geq 0,$$

where  $\alpha, \beta > 0$ . It can be easily proved that it belongs to the exponential family, with  $\tau = 1/\alpha$ ,  $\theta = -\beta/\alpha$  and  $a(\theta) = \ln(-1/\theta) = \ln(\alpha/\beta)$ .

Now assume

$$Y_{ij}|\theta_{ij}, \tau \stackrel{\text{iid}}{\sim} p(y_{ij}|\theta_{ij}, \tau), \quad \text{where } j = 1, 2, \dots, n_i, \quad (2.2)$$

so that each observation has its own canonical parameter  $\theta_{ij}$  and all observations share the same dispersion parameter  $\tau$ .

In the generalized linear models  $\theta_{ij}$  is linked to the covariates by

$$h(\theta_{ij}) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta},$$

where  $h(\cdot)$  is monotone and is called  $\theta$ -link,  $\eta_{ij}$  is named linear predictor, and  $\mathbf{x}_{ij}$  is the row  $j$  of the matrix  $X_i$ .

In a GLMM we consider the random effects  $\boldsymbol{\alpha}_i \sim N_v(\mathbf{0}, \Lambda)$  in addition to the fixed ones

$$h(\theta_{ij}) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\alpha}_i,$$

where  $\mathbf{z}_{ij}$  is the row  $j$  of the matrix  $Z_i$ , to capture extra-variability of the unit  $i$ .

Now we can write  $f(y_{ij}|\theta_{ij}, \tau)$  as a function of  $\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \tau$ :

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \tau) = \exp\{\tau[y_{ij}h^{-1}(\eta_{ij}) - a(h^{-1}(\eta_{ij}))] + c(y_{ij}, \tau)\} \quad (2.3)$$

Notice that, given  $\boldsymbol{\alpha}_i$ , the repeated observations of the unit  $i$  are independent thanks to (2.2). Thus, the likelihood for  $N$  statistical units is

$$L(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{i=1}^N \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \tau), \quad (2.4)$$

where  $\mathbf{y}' = (y_{11}, \dots, y_{1n_1}, \dots, y_{N1}, \dots, y_{Nn_N})$  and  $\boldsymbol{\alpha}' = (\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \dots, \boldsymbol{\alpha}'_N)$ .

Hence the vector of the parameters is  $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau, \sigma^2, \Lambda)$ .

## 2.3 Bayesian approaches to the generalized linear mixed models

In a Bayesian parametric framework the parameters  $\boldsymbol{\beta}$  and  $\Lambda$  are random. In this work we will consider a model with just one-dimensional random effects, i.e.  $v = 1$ , and so, in place of the multivariate  $\Lambda$ , we will consider the univariate  $\lambda$ .



In this case, the likelihood is as in (2.4), with  $f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \tau)$  as in (2.3). The "standard" parametric prior for  $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau, \sigma^2, \lambda)$  is as in Zeger and Karim [38]

$$\begin{aligned}\boldsymbol{\beta} &\sim N_p(\nu_0, \Sigma_0) \\ \alpha_1, \alpha_2, \dots, \alpha_N | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\ \lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right),\end{aligned}$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are independent, and  $\tau, \sigma^2$  are fixed.

We recall that a random variable  $Y \sim \text{InvGamma}(\alpha, \beta)$  (Inverse-gamma-distributed with shape  $\alpha > 0$  and scale  $\beta > 0$ ) has density function

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\frac{\beta}{y}\right), \quad y \geq 0$$

Hence  $\alpha_1, \dots, \alpha_N$  are exchangeable for Theorem 1. The exchangeability assumption for the random effects is appropriate since the  $N$  statistical units can be seen as a finite sample without replacement from a larger population of units. Moreover, exchangeability enables to "borrow strength" across random effects by linking observations through a covariance model, and so it improves individual random effect estimates.

Kleinman and Ibrahim [23] relaxed the normality assumption and considered  $\alpha_1, \alpha_2, \dots, \alpha_N | P \stackrel{\text{iid}}{\sim} P$  and  $P \sim DP(\alpha)$ . Moreover they introduce a prior *Gamma* to the dispersion parameter  $\tau$ . The prior specifications for the parameters are

$$\begin{aligned}\boldsymbol{\beta} &\sim N_p(\nu_0, \Sigma_0) \\ \alpha_1, \alpha_2, \dots, \alpha_N | P &\stackrel{\text{iid}}{\sim} P \\ P &\sim DP(aN(0, \lambda)) \\ \lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \\ \tau &\sim \text{Gamma}(a_0, b_0)\end{aligned}$$

Relaxing the normality assumption may better express our uncertainty about the true distribution of the random effects. We recall that the random effects describe a latent and unknown structure between the statistical units, and hence in general we have little prior information about their distribution.

Of course it is important to accurately model the distribution of the random effects in particular when we are interested in the prediction for a future observation from a given subject, since the posterior distribution of the random effects can be affected by its prior distribution. For example the normal random effects model can perform poorly when the random effects have a multi-modal distribution.

Moreover it is desirable to relax the assumption of normality when we are interested in doing inference about the distribution of the random effects itself.

## Chapter 3

# Application to accelerated failure time models

### 3.1 Survival analysis

Survival analysis is a branch of statistics focused on analyzing time-to-event data. These modeling techniques have application in engineering, medicine, public health, economics among others. We mention Kalbfleisch and Prentice [20] for a comprehensive presentation of the main models and methods in survival analysis.

The target of a survival regression study is to understand the dependence between time-to-event (often called failure time) and some explanatory variables. For example, the time to recovery of a patient can be related to the age, the sex or a specific treatment.

A difficulty that frequently arises in trials having time-to-event endpoints is that a fraction of the subjects could not "fail" at the end of the study. For these subjects it is only known that the true time-to-event is to the right of the conclusion of the trial. These times are called right-censored data.

Often the interest in survival analysis lies in the survival function and the hazard function. Let  $T$  be an absolutely continuous random variable representing the failure time of a subject and let  $f(\cdot)$  denotes the probability density of  $T$ .

The survival function  $S(t)$  represents the probability that the individual time-to-event is greater than  $t$ :

$$S(t) := \mathbb{P}(T > t), \quad t > 0,$$

while the hazard function  $h(t)$  is the instantaneous rate of failure upon time  $t$ :

$$h(t) := \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t > 0.$$

Frequentist techniques in survival analysis are based on the estimation of the survival function or the hazard function. In particular we mention the Kaplan Meier estimator,

that estimates the survival function  $S(\cdot)$  through a nonparametric likelihood approach, and the Cox proportional hazard model. Under the latter model different subjects have hazard functions that are proportional to one another, and the covariates have a multiplicative effect on the hazard function  $h(\cdot)$ .

In a more recent model called *accelerated failure time model* (AFT) the covariates influence the survival time  $T$  through the following relationship

$$\log(T) = X\boldsymbol{\beta} + W$$

or equivalently

$$T = \exp(X\boldsymbol{\beta})V, \quad V = e^W > 0,$$

where  $X$  denotes the matrix of covariates, and  $\boldsymbol{\beta}$  is the vector of regression parameters. Observe that the covariates  $X$  may *accelerate* or *decelerate* the time to failure. The AFT models can be framed as generalized linear models for the survival time  $T$ , where  $T$  can assume a variety of distributions on  $\mathbb{R}^+$  (Weibull, log-normal, gamma, ...).

### 3.2 Accelerated life test on NASA pressure vessels

The analysis of accelerated life tests (ALTs) is very important in reliability. We recall that in engineering applications survival analysis is generally called reliability. ALT testing involves acceleration of failures with the purpose of predicting the life characteristics of the product at normal use conditions.

Here we analyze an ALT test on NASA pressure vessels, which are critical components of the Space Shuttle. In particular we model the failure times via a semi-parametric Bayesian AFT model, and provide posterior estimates of the regression parameters and credibility intervals of some given quantiles of the failure-time distribution as well.

The application is based on a dataset of 154 lifetimes of pressure vessels wrapped with Kevlar yarn. The fiber comes from 8 different spools and 5 levels of pressure stress are used. All the vessels at the lowest level of stress of 17.2MPa and 11 vessels out of 21 at stress level 23.4MPa are censored at 41000 hours. Further details on the pressure vessels are given in Section 3.2.1. Several authors made a statistical analysis of this dataset, but nobody included the 46 vessels at lowest level of stress in his own analysis. Although we think that the 46 vessels should be included in the analysis via a hierarchical model, we decide to omit them in order to compare our results to those of other authors; therefore we will consider 108 times (97 failure-times and 11 right-censored), from 8 different spools and under 4 levels of pressure (the data-set is provided in Appendix C).

Glaser (1983) [16], Gerstle and Kunz (1984) [14] and Crowder et al. (1991) [6] used frequentist approaches to the problem and considered the spool effect as fixed. Feiveson and Kulkarni (2000) [10] (who omitted the spool 7 from their analysis, see Section 3.2.1)

emphasize the necessity of treating the spool effect as random. Leon et al. (2006) [24] considered a Bayesian Weibull regression with random effects, and Argiento et al. (2010) [2] proposed a semiparametric Bayesian Weibull regression model.

The results of those authors are quite sensible to the modeling assumptions: the authors that considered the spool effect as fixed concluded that the risk was minimal; while the authors that treated the spool as a random-effect concluded that there is not enough information in the data-set to make statements about the reliability of the pressure vessels.

In our point of view, a model with spool effect treated as fixed can be used only to make prediction about the life-time of a new pressure vessel wrapped with Kevlar from either one of the 8 known spools or from an "average" of them. Unfortunately, we do not know which spool is used to wrap the pressure vessels of the Space Shuttle; in fact, the vessels may be wound from one of the 8 known spool, or from a spool selected at random from the population of spools. Hence, we will treat the spool affect as random to predict properly the failure-time of a new pressure vessel wrapped with Kevlar from an unknown spool.

### 3.2.1 NASA pressure vessels and Kevlar fibres

A pressure vessel is a closed container designed to hold gases or liquids at high pressure. Pressure vessels wrapped with composite materials have been largely used. In fact, in applications with one short pressure cycle, such as rocket motors, composite vessels offer considerable weight savings over steel or titanium alternatives.

The short-term static-burst behavior of such vessels is quite predictable. On the contrary the response of composite vessels to cyclic or sustained pressurization is not predictable. A study was undertaken in 1976 to determine what constitutes a safe design stress level for Kevlar 49 Epoxy under constant load for a specified time. The data comes from Kevlar 49 Epoxy creep-rupture experiments conducted during five years at U.S. Department of Energy Lawrence Livermore National Laboratories on scaled-down replicates of NASA pressure vessels.

Each vessel consisted of a 112mm internal diameter aluminum liner, 1mm thick, over-wrapped with 1.1mm of Kevlar 49 yarn wetted during winding with epoxy (see Figure 3.1). The liner is made principally from aluminum hemispheres, electron-beam welded at the equator. Although the liner is nearly as thick as the composite, it is much weaker and therefore contributes little to the burst strength of the vessel.

Moreover, we have further information about the microscopic characteristics of a single fiber and a single yarn of spool 2 and 7, respectively. In Table 3.1 we report the results of mechanical tests that were made on both a single fiber and a single yarn of these spools. In our analysis in Sections 4.1.1 and 4.2.1, we will conclude that spool 7 performs worse than spool 2. From Table 3.1, the low performance of Spool 7 can be explained by the

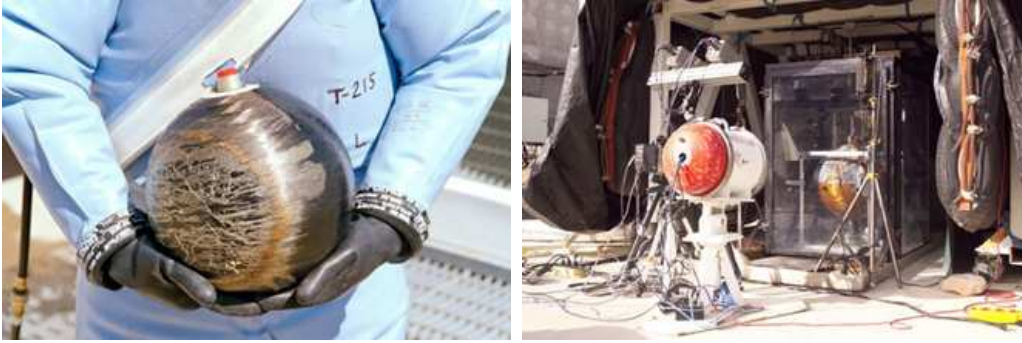


Figure 3.1: Photos of the NASA pressure vessels from the 2009 Composite Pressure Vessel and Structure Summit [5].

microscopic characteristics of the yarn; in fact it has only 127 filaments instead of the nominal 267 and the diameter of the fiber is 17mm (instead of 11mm). We guess that the Spool 2 has better performance since its properties are closer to the nominal ones. While Feiveson and Kulkarni [10] omitted Spool 7 from their analysis since it does not respect the nominal properties, we decided to keep it here for two reasons. First, we are interested in the identification of other spools with performance similar to Spool 7. Secondly, since the fibers and the yarns of spools are not tested one-by-one, we will over-estimate the performance of a new pressure vessel from an unknown spool if we exclude the "bad" spools.

	Nominal	Spool 2	Spool 7
Fiber strength (GPa)	2.86	3.33	3.31
Yarn failure load (Kg)	7.3	6.6	6.0
Fiber diameter ( $\mu\text{m}$ )	11.0	11.5	17.0
Fibers per strand of yarn	267	263	127
Yarn strength (GPa)	2.76	2.59	2.02

Table 3.1: Mechanical properties of a single fiber and a single yarn of Kevlar

### 3.3 Accelerated life models for Kevlar fiber failure times

First we introduce the parametric AFT model considered in Leon et al. [24], then we propose two different semi-parametric AFT models that generalize the parametric one. In the former the linear predictor is handled parametrically and the errors are modelled as a Weibull mixture of a NGG process. In the latter the random effects normality assumption is relaxed and they are treated non-parametrically, while the error is parametric.

Finally we briefly present a semi-parametric AFT model with non-parametric random effects and with gamma-distributed errors. We will fit this model using *DPpackage* package [19] for Bayesian nonparametric modelling in the programming software *R* [33].

### 3.3.1 Parametric model

Leon et al. [24] fitted a parametric Bayesian AFT mixed model assuming Weibull survival times. They considered the stress as a fixed effect, and the spool as a random one. As mentioned in Section 3.1, we can write the model in additive form as

$$\begin{aligned}\log T &= \beta_0 + \beta_1 \log(x_s) + \boldsymbol{\alpha}' \mathbf{x} + \frac{W}{\theta} \\ W &\sim \text{Gumbel}(0, 1)\end{aligned}\tag{3.1}$$

or, equivalently, in multiplicative form

$$\begin{aligned}T &= \exp\{\beta_0 + \beta_1 \log(x_s) + \boldsymbol{\alpha}' \mathbf{x}\} V \\ V &\sim \text{Weibull}(\theta, 1),\end{aligned}\tag{3.2}$$

where  $x_s = \text{stress}$ ,  $\mathbf{x} = (x_1, \dots, x_J)$ ,  $x_j$  is a binary covariate to identify the spool  $j$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)$ ,  $j = 1, \dots, J$ , and  $J = 8$  is the number of spools.

In both model representations the error has a standard distribution. In the additive model (3.1)  $\theta$  is a scale parameter, while in the multiplicative model (3.2) is a shape parameter. We recall that the probability density function of the standard Gumbel r.v.  $W$  is

$$f(w) = \exp\{w - \exp(w)\},$$

with its survival function

$$S(w) = \exp\{-\exp(w)\}.$$

While the probability density function of the Weibull r.v.  $V$  is

$$f(v) = \frac{\theta}{\lambda} \left(\frac{v}{\lambda}\right)^{\theta-1} \exp\left\{-\left(\frac{v}{\lambda}\right)^\theta\right\}, \quad v \geq 0,$$

with its survival function

$$S(v) = \exp\left\{-\left(\frac{v}{\lambda}\right)^\theta\right\}, \quad v \geq 0,$$

where  $\theta > 0$  is the shape parameter and  $\lambda > 0$  is the scale one. We recall that the Weibull r.v. has the following scaling property:

$$V \sim \text{Weibull}(\theta, \lambda) \Leftrightarrow cV \sim \text{Weibull}(\theta, c\lambda),\tag{3.3}$$

where  $c > 0$ .

We also notice that neither the Weibull nor the Gumbel distribution belong to the exponential family and so (3.1) and (3.2) are not exactly GLMMs. However, they can be considered as straightforward generalizations.

Leon et al. [24] assumes the following model

$$V_i|\theta \stackrel{\text{iid}}{\sim} Weibull(\theta, 1), \quad i = 1, \dots, n, \quad (3.4)$$

with priors

$$\begin{aligned} \beta_0 &\sim N(0, \sigma_0^2) \\ \beta_1 &\sim N(0, \sigma_1^2) \\ \alpha_1, \dots, \alpha_J | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\ \lambda &\sim InvGamma\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \\ \theta &\sim gamma(a_0, b_0), \\ \beta_0, \beta_1, \boldsymbol{\alpha}, \theta &\text{ independent.} \end{aligned} \quad (3.5)$$

Here we have  $n = 108$  observations, where  $m = 11$  right censored times  $T_{cens}$  and  $(n - m) = 97$  failure times.

From (3.4), (3.5) and the scaling property (3.3) it follows that

$$T_i|\beta_0, \beta_1, x_{s,i}, \mathbf{x}_i, \boldsymbol{\alpha}, \theta \stackrel{\text{ind}}{\sim} Weibull(\theta, \exp\{\beta_0 + \beta_1 \log(x_{s,i}) + \boldsymbol{\alpha}' \mathbf{x}_i\}).$$

This is an AFT model and, as we mentioned in Section 3.1, the linear predictor

$$\eta_i := \exp\{\beta_0 + \beta_1 \log(x_{s,i}) + \boldsymbol{\alpha}' \mathbf{x}_i\} \quad (3.6)$$

accelerates or decelerates the failure time  $T_i$ . Instead of  $\boldsymbol{\alpha}' \mathbf{x}_i$ , we will generally use  $\alpha_{k[i]}$ , where  $k[i] = 1, \dots, J$ , to identify the spool of observation  $i$ .

### 3.3.2 Nonparametric error model

In (3.4) and (3.5), all observations share the same shape parameter  $\theta$ . We relax this hypothesis by letting the error distribution to be a shape mixture of Weibull densities, where the mixing measure is a NGG. Since we think that all observations could not necessarily share the same shape parameter. In particular, we suspect that the spools that do not respect the nominal conditions like Spool 7 (see Table 3.1) may have a different shape parameter.

Now the model is

$$V_i|\theta_i \stackrel{\text{ind}}{\sim} Weibull(\theta_i, 1), \quad i = 1, \dots, n, \quad (3.7)$$

under the prior

$$\begin{aligned}
\beta_0 &\sim N(0, \sigma_0^2) \\
\beta_1 &\sim N(0, \sigma_1^2) \\
\alpha_1, \dots, \alpha_J | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\
\lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \\
\theta_1, \dots, \theta_n | P &\sim P \\
P &\sim \text{NGG}(\sigma, \omega, \kappa, P_0) \\
P_0 &\sim \text{Gamma}(a_0, b_0), \\
\beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta} &\text{ independent.}
\end{aligned} \tag{3.8}$$

Here we have  $m$  right-censored times and  $(n - m)$  failure times. This means that

$$T_i | \beta_0, \beta_1, x_{s,i}, \alpha_{k[i]}, \theta_i \stackrel{\text{iid}}{\sim} \text{Weibull}(\theta_i, \exp\{\beta_0 + \beta_1 \log(x_{s,i}) + \alpha_{k[i]}\}). \tag{3.9}$$

Observe that, conditionally on the r.p.m.  $P$ , the error terms are i.i.d. according to the nonparametric mixture, which represents a generalization of the well known Dirichlet mixture process model (see Lo [28] and Ferguson [11]).

We notice that the grouping of  $\theta_i$  is random and inferred a posteriori from the data. Moreover, conditionally on the mixing measure  $P$ , the error terms are such that

$$V_i | P \stackrel{\text{iid}}{\sim} \int_{\mathbb{R}^+} k(v; \theta, 1) P(d\theta) \quad i = 1, \dots, n,$$

where  $k(\cdot; \theta, 1)$  denotes the kernel density of a Weibull distribution with shape parameter  $\theta$  and scale equal to 1.

We are mainly interested in the predictive distribution of the life time  $T_{n+1, \text{new}}$  of a new pressure vessel wrapped with Kevlar fiber from a random unknown spool  $\alpha_{\text{new}}$  and subject to a given stress level  $x_s$ . The predictive probability density distribution  $f_{T_{n+1, \text{new}}}(\cdot | \mathbf{T})$  is

$$\begin{aligned}
f_{T_{n+1, \text{new}}}(t | \mathbf{T}) &= \mathbb{E}\left[\int_{\mathbb{R}^+} k(t; \theta, \exp\{\beta_0 + \beta_1 \log(x_s) + \alpha_{\text{new}}\}) P(d\theta) | \mathbf{T}\right] \\
&= \int \left\{ \int_{\mathbb{R}^+} k(t; \theta, \exp\{\beta_0 + \beta_1 \log(x_s) + \alpha_{\text{new}}\}) P(d\theta) \right\} \times \\
&\quad \mathcal{L}(d\beta_0, d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\lambda, dP | \mathbf{T}),
\end{aligned} \tag{3.10}$$

where  $t > 0$  and  $\mathbf{T} = (T_1, \dots, T_n)$ .

The posterior distribution

$$\mathcal{L}(d\beta_0, d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\lambda, dP | \mathbf{T}) \tag{3.11}$$

in (3.10) does not have any tractable analytic expression. So the predictive distribution (3.10) must be computed by Markov Chain Monte Carlo integration. If we sample a large



enough Markov sequence  $\{\beta_0^{(b)}, \beta_1^{(b)}, \alpha_{\text{new}}^{(b)}, P^{(b)}\}_{b=1}^B$  whose limit distribution is (3.11), we will estimate  $f_{T_{n+1}, \text{new}}(\cdot | \mathbf{T})$  by the ergodic mean

$$\frac{1}{B} \sum_{b=1}^B \int_{\mathbb{R}^+} k(t; \theta, \exp\{\beta_0^{(b)} + \beta_1^{(b)} \log(x_s) + \alpha_{\text{new}}^{(b)}\}) P^{(b)}(d\theta).$$

By (1.6), the MCMC estimate of the predictive distribution is

$$\frac{1}{B} \sum_{b=1}^B \sum_{i=1}^{\infty} k(t; \tau_i^{(b)}, \exp\{\beta_0^{(b)} + \beta_1^{(b)} \log(x_s) + \alpha_{\text{new}}^{(b)}\}) p_i^{(b)}.$$

Of course, the predictive distribution of the life  $T_{n+1, j}$  of a new pressure vessel with Kevlar fiber from a given spool  $j$  can be computed as

$$f_{T_{n+1, j}}(t | \mathbf{T}) = \int \left\{ \int_{\mathbb{R}^+} k(t; \theta, \exp\{\beta_0 + \beta_1 \log(x_s) + \alpha_j\}) P(d\theta) \right\} \times \\ \mathcal{L}(d\beta_0, d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\lambda, dP | \mathbf{T}),$$

and it can be estimated by

$$\frac{1}{B} \sum_{b=1}^B \sum_{i=1}^{\infty} k(t; \tau_i^{(b)}, \exp\{\beta_0^{(b)} + \beta_1^{(b)} \log(x_s) + \alpha_j^{(b)}\}) p_i^{(b)}.$$

To sample from the posterior distribution (3.11) we build a Gibbs sampler with limit distribution equal to the posterior. We augment the state space with an auxiliary variable  $U$  to be able to sample from all the full-conditional posterior distributions. We have introduced this latent variable in (1.13) to provide a tractable expression of the posterior distribution of a NGG: now it will be used to sample from the full-conditional of  $P$ .

The posterior distribution (3.11) is a function of  $\mathbf{T} = (T_1, \dots, T_n)$ , while we are given only  $T_1, \dots, T_{n-m}$  failure times and  $m$  right-censored times  $T_{\text{cens}}$ . Hence we must impute  $T_{n-m+1}, \dots, T_n$  and include them into the state space of the Markov Chain. The imputed times  $T_i$ , for  $i = n - m + 1, \dots, n$ , can be obtained easily from the following truncated Weibull distribution

$$\mathbb{P}(T_i > t | T_i > T_{\text{cens}}, x_{s, i}, \beta_0, \beta_1, \alpha_{k[i]}, \theta_i) = \begin{cases} \frac{\mathbb{P}(T_i > t | x_{s, i}, \beta_0, \beta_1, \alpha_{k[i]}, \theta_i)}{\mathbb{P}(T_i > T_{\text{cens}} | x_{s, i}, \beta_0, \beta_1, \alpha_{k[i]}, \theta_i)} & \text{if } t > T_{\text{cens}} \\ 1 & \text{if } t \leq T_{\text{cens}} \end{cases}$$

In conclusion, we will sequentially draw from the following full-conditionals

$$\begin{aligned}
& \mathcal{L}(d\beta_0|\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T}) \\
& \mathcal{L}(d\beta_1|\beta_0, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T}) \\
& \mathcal{L}(d\alpha_j|\beta_0, \beta_1, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T}), \quad j = 1, \dots, J \\
& \mathcal{L}(d\alpha_{\text{new}}|\beta_0\beta_1, \boldsymbol{\alpha}, \lambda, P, U, \mathbf{T}) \\
& \mathcal{L}(d\lambda|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, P, U, \mathbf{T}) \\
& \mathcal{L}(dP|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, U, \mathbf{T}) \\
& \mathcal{L}(dU|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \mathbf{T}) \\
& \mathcal{L}(dT_i|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T}_{-i}), \quad i = n - m + 1, n - m + 2, \dots, n.
\end{aligned} \tag{3.12}$$

As usual,  $\boldsymbol{\alpha}_{-j}$  indicates the vector  $\boldsymbol{\alpha}$  without the  $j$ -th component, i.e.

$\boldsymbol{\alpha}_{-j} = (\alpha_1, \dots, \alpha_{j-1}, \alpha_{j+1}, \dots, \alpha_J)$ . Similarly  $\mathbf{T}_{-i}$  is the vector  $\mathbf{T}$  without  $T_i$ .

Observe that at each iteration of the algorithm we simulate a trajectory from the posterior distribution of  $P$ . We need complete trajectories of  $P$  to provide credibility intervals of any given quantiles of the failure-time distribution. In fact, if we sampled only from the full-conditionals of  $\theta_i$ 's and not from that of  $P$ , we would lose one level of the hierarchy, and hence we would underestimate the variability of the marginal error  $V$ .

The algorithms that simulate complete trajectories from the posterior distribution of  $P$  are very recent in the literature and were introduced by [17]; of course, they are more difficult to implement and more time consuming with respect to the popular *Polya Urn Scheme* algorithms that integrate out the process  $P$  and sample only from the full-conditionals of  $\theta_i$ 's. However they provide more information on the structure of the model. We recall that the *Polya Urn Scheme* algorithms are based on the idea of the *Blackwell-MacQueen urn scheme* (1.1) and were introduced by Escobar [8], and Escobar and West [9] for Bayesian density estimation.

It can be proved that the full-conditional of  $P$  in (3.12) is independent of  $\mathbf{T}$  and of the linear predictor  $\eta_i$  (3.6), and so

$$\mathcal{L}(dP|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, \boldsymbol{\theta}, U, \mathbf{T}) = \mathcal{L}(dP|\boldsymbol{\theta}, U) \tag{3.13}$$

By (1.13), we know that (3.13) is a mixture of a NGG process and a discrete r.p.m. with finite fixed points. Hence the posterior law of  $P$  admits an infinite series representation, but we will sample only a finite series using a stopping criterion as in Argiento et al. [1]. Although in this way we obtain only approximated trajectories of  $P$ , this stopping criterion guarantees the convergence for functionals of  $P$ . We mention that Walker [37] and Kalli et al. [21] recently proposed methods to sample exactly from the posterior of  $P$  by introducing a latent variable.

The explicit expressions of all full-conditionals and further information on the algorithm are provided in Appendix A.

### 3.3.3 Nonparametric random effects model

As we will see in the results Chapter 4, treating the error distribution as a shape mixture of Weibull densities does not improve remarkably the predictive capability of the parametric model (3.4)-(3.5). Therefore we tried to generalize it in a different manner, i.e. relaxing the normality assumption for the random-effects parameters; in fact, we allow now the distribution for the spool effect to be multi-modal, in order to not over-estimate its variance as before. Here we handle the random effects nonparametrically, while the shape parameter of the error is parametrically distributed as in Leon et al. (3.4). The model we propose in this Section is similar to those in Kleinmann and Ibrahim [23], where the authors used a Dirichlet prior for the random effects, while here we use a NGG prior. As mentioned in Section 1.4, the NGG process prior is ruled by two parameters (instead of one) controlling the clustering mechanism; so the NGG is more flexible than the Dirichlet process. The model is

$$V_i|\theta \stackrel{\text{iid}}{\sim} Weibull(\theta, 1) \quad i = 1, \dots, n, \quad (3.14)$$

and the prior specifications for the parameters are

$$\begin{aligned} \theta &\sim Gamma(a_0, b_0) \\ \beta_1 &\sim N(0, \sigma_1^2) \\ \alpha_1, \dots, \alpha_J | P &\stackrel{\text{iid}}{\sim} P \\ P &\sim NGG(\sigma, \omega, \kappa, P_0) \\ P_0 | \mu, \lambda &\sim N(\mu, \lambda) \\ \lambda &\sim InvGamma\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \\ \mu &\sim N(0, \sigma_0^2), \\ \beta_1, \alpha, \theta &\text{ independent.} \end{aligned} \quad (3.15)$$

For (3.14), (3.15) and the scaling property (3.3), it follows that the failure time  $T_i$  has conditional distribution

$$T_i | \beta_1, x_{s,i}, \alpha_{k[i]}, \theta \stackrel{\text{ind}}{\sim} Weibull(\theta, \exp\{\beta_1 \log(x_{s,i}) + \alpha_{k[i]}\}).$$

In (3.9) we consider the mean effect  $\beta_0$  and the random effects  $\alpha_j$  separately, for  $j = 1, \dots, J$ . However, in the corresponding Markov Chain trajectories we observed a strong correlation between the mean effect  $\beta_0$  and each random effect  $\alpha_j$ , and so we needed a long Gibbs Sampler to converge to the limit distribution. Here we change parameterization to speed convergence: we eliminate the mean effect and we add a prior on the mean of the base-line  $P_0$  of the random effects. We cannot distinguish the difference between the mean effect and the spool effect anymore, but this is not a limitation since we are interested

in the predictive distribution rather than the posterior estimates. Recently Li et al. [25] proposed a posterior technique to separate the mean effect from the random one when a Dirichlet process prior is used.

As already mentioned in Section 3.3.2, we are interested in the predictive distribution of the life time  $T_{n+1,\text{new}}$  of a new pressure vessel wrapped with Kevlar fiber from a random unknown spool  $\alpha_{\text{new}}$  and subject to a given stress level  $x_s$ . The predictive probability density distribution  $f_{T_{n+1,\text{new}}}(\cdot|\mathbf{T})$  is

$$f_{T_{n+1,\text{new}}}(t|\mathbf{T}) = \int k(t; \theta, \exp\{\beta_1 \log(x_s) + \alpha_{\text{new}}\}) \mathcal{L}(d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\theta, d\mu, d\lambda|\mathbf{T}), \quad (3.16)$$

where  $t > 0$  and  $\mathbf{T} = (T_1, \dots, T_n)$ . If we were interested in the predictive distribution of the life-time  $T_{n+1,j}$  of a new pressure vessel with Kevlar from a given spool  $j$ , we could simply replace  $\alpha_{\text{new}}$  with  $\alpha_j$  in the linear predictor in (3.16).

Since there is no manageable analytic formula for the posterior distribution

$$\mathcal{L}(d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\theta, d\mu, d\lambda|\mathbf{T}), \quad (3.17)$$

we build a Gibbs sampler that converges to it. Then we will estimate the predictive probability density distribution  $f_{T_{n+1,\text{new}}}(\cdot|\mathbf{T})$  by the ergodic mean

$$\frac{1}{B} \sum_{b=1}^B k(t; \theta^{(b)}, \exp\{\beta_1^{(b)} \log(x_s) + \alpha_{\text{new}}^{(b)}\}),$$

where  $\{\beta_1^{(b)}, \alpha_{\text{new}}^{(b)}, \theta^{(b)}\}_{b=1}^B$  is a Markov sequence from the Gibbs sampler whose limit distribution is (3.17).

As in the previous Section, we must impute the  $m$  right-censored times  $T_{\text{cens}}$  and include them into the Markov Chain. It can be easily proved that they have truncated Weibull distribution (see 3.3.2).

The full-conditionals of the Gibbs sampler are

$$\begin{aligned} & \mathcal{L}(d\beta_1|\boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta, \mathbf{T}) \\ & \mathcal{L}(d\alpha_j|\beta_1, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \mu, \lambda, \theta, \mathbf{T}), \quad j = 1, \dots, J \\ & \mathcal{L}(d\alpha_{\text{new}}|\beta_1, \boldsymbol{\alpha}, \mu, \lambda, \theta, \mathbf{T}) \\ & \mathcal{L}(d\mu|\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, \theta, \mathbf{T}) \\ & \mathcal{L}(d\lambda|\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \theta, \mathbf{T}) \\ & \mathcal{L}(d\theta|\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \mathbf{T}) \\ & \mathcal{L}(dT_i|\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta, \mathbf{T}_{-i}), \quad i = n - m + 1, n - m + 2, \dots, n. \end{aligned} \quad (3.18)$$

Observe that we do not need to sample any trajectory of  $P$  to provide credibility intervals of some given quantiles of the failure-time distribution for a new spool or a given spool

$j$ . In fact  $\alpha_{jS}$  and  $\alpha_{\text{new}}$  are part of the linear predictor, and the linear predictor has only a multiplicative effect as showed in (3.6). Hence, here we do not simulate trajectories of the process  $P$ , but we will use a Polya-urn scheme algorithm to sample the full-conditionals of  $\alpha_j$ 's.

This Gibbs sampler in (3.18) is computationally less heavy than (3.12), since it provides single values  $\alpha_j^{(b)}$ 's and  $\alpha_{\text{new}}^{(b)}$ , and not a complete trajectories  $P^{(b)}$ , at each iteration  $b$ . However it is not trivial to sample from all the full-conditionals. In particular to sample from the full-conditionals of  $\alpha_j$ s we implement the so-called Algorithm 2 of Neal [30] for non-conjugate priors; while the full-conditionals of  $\beta_1$  and  $\theta$  are log-concave and so we use an adaptive rejection sampling method [36]. In Appendix B we provide further information about the sampling methods used for each full-conditional of the Gibbs sampler for (3.14)-(3.15).

### 3.3.4 DPpackage: gamma-distributed error model

*DPpackage* [19] is a package for the modelling programming software *R* [33] that offers several functions to fit Bayesian nonparametric and semiparametric models. In particular the function *DPglm* fits GLMMs with nonparametric random effects as in Kleinmann and Ibrahim [23] (see Section 2.2 to a brief introduction to GLMM).

As mentioned in Section 3.3.1, the Weibull distribution does not belong to the exponential family, and so (3.2) is not exactly a GLMM. Since the function *DPglm* fits models with error from the exponential family, we replaced the Weibull with a Gamma distribution and used the package to "explore" the data-set. We recall that the Weibull distribution is popular in AFT models since it has a natural interpretation of the shape and scale parameters. The Gamma distribution has a similar density function, and hence we expect that our results will not be too much biased by this error distribution choice. Since the function *DPglm* does not take into account right-censored data, we must discard the  $m$  right-censored times. The model we fit using *DPglm* has the following likelihood

$$V_i | \theta \stackrel{\text{iid}}{\sim} \text{Gamma}\left(\frac{1}{\theta}, \frac{1}{\theta}\right) \quad i = 1, \dots, (n - m), \quad (3.19)$$

and the these prior specifications for the parameters

$$\begin{aligned}
\theta &\sim \text{Gamma}(a_0, b_0) \\
\beta_1 &\sim N(0, \sigma_1^2) \\
\alpha_1, \dots, \alpha_J | P &\stackrel{\text{iid}}{\sim} P \\
P &\sim \text{DP}(aP_0) \\
a &\sim \text{Gamma}(a_1, b_1) \\
P_0 | \mu, \lambda &\sim N(\mu, \lambda) \\
\lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \\
\mu &\sim N(0, \sigma_0^2), \\
\beta_1, \boldsymbol{\alpha}, \theta &\text{ independent.}
\end{aligned} \tag{3.20}$$

We recall that the Gamma r.v. also has the scaling property:

$$V \sim \text{Gamma}(\theta, \lambda) \Leftrightarrow cV \sim \text{Gamma}\left(\theta, \frac{\lambda}{c}\right), \tag{3.21}$$

where  $c > 0$ .

Hence, for (3.19), (3.20) and the scaling property (3.21), it follows that the failure time  $T_i$  has distribution

$$T_i | \beta_1, x_{s,i}, \alpha_{k[i]}, \theta \stackrel{\text{ind}}{\sim} \text{Gamma}\left(\frac{1}{\theta}, \frac{1}{\theta} \exp\{-\beta_1 \log(x_{s,i}) - \alpha_{k[i]}\}\right).$$

Notice that  $\mathbb{E}[V|\theta] = 1$  and  $\text{Var}[V|\theta] = \theta$ , and hence  $\theta$  is the dispersion parameter, while the linear predictor

$$\eta_i := \exp\{\beta_1 \log(x_{s,i}) + \alpha_{k[i]}\}$$

is a multiplicative factor that accelerates or decelerates the failure time  $T_i$ .

Observe that in (3.20), we assume a Dirichlet process prior in each of the random-effects parameters  $\alpha_1, \dots, \alpha_J$ , and the mass parameter  $a$  of the Dirichlet process is given a Gamma prior.

The function *DPglm* provides a sample from the posterior distribution

$$\mathcal{L}(d\beta_1, d\boldsymbol{\alpha}, d\alpha_{\text{new}}, d\theta, da, d\mu, d\lambda | \mathbf{T}),$$

using a *Polya-Urn scheme* algorithm. In Section 4.3.3, we give further information on the input arguments and output elements of *DPglm*.

# Chapter 4

## Results

In this chapter we analyze the results of our semi-parametric models, and compare them to those obtained by Leon et al. [24]. We report interval estimates of the quantiles of the predictive distributions for a new spool, and for Spool 2 and 7. It is interesting to consider these latter spools, since we have information about the microscopic characteristics of the yarn and the fiber of these spools. In particular we provide interval estimates for the first percentile of the failure time distribution at stress level 23.4MPa, the lowest value in the dataset, and for the median time when stress is 22.5MPa (extrapolated). Leon et al. [24] chose the latter stress level because it is lower than those in the experiment, but close enough to provide reasonable estimate. We recall that in ALT tests, the stress levels are higher than normal use condition to accelerate the failure times.

Before presenting our results, we report the numerical values of the priors' hyper-parameters in the parametric model of Leon et al. (3.4)-(3.5), and a summary of their results.

The priors specifications are

$$\begin{aligned}\beta_0 &\sim N(0, \sigma_0^2 = 0.001) \\ \beta_1 &\sim N(0, \sigma_1^2 = 0.001) \\ \alpha_1, \dots, \alpha_J | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\ \lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2} = 0.001, \frac{\tau_2}{2} = 0.001\right) \\ \theta &\sim \text{Gamma}(a_0 = 1, b_0 = 0.2).\end{aligned}\tag{4.1}$$

Notice that, if  $\tau := \tau_1 = \tau_2$ , each  $\alpha_j$  has a marginal t-student prior with  $\tau$  degrees of freedom. Leon et al. [24] consider  $\lambda$  in (4.1) to have a vague prior, since a t-student with  $\tau = 0.002$  degrees of freedom has heavy tails. On the contrary we think that such a prior is greatly informative, since it puts most of its mass in neighborhood of zero and infinity.

Moreover they chose the hyperparameters of  $\theta$  on the basis of some engineering arguments about the shape parameter of the Weibull distribution. Indeed, if  $\theta > 1$  the failure rate increases with time, otherwise, if  $\theta < 1$  it decreases over time. Although those prior beliefs are reasonable, we think that they should also take into account the marginal distribution of the error  $V$ . Marginally, there are no analytic expressions of the density distribution function of the error; but we can provide numerical estimation of its density function (see Figure 4.1a), and analytic expressions of the first to moments of  $\log(V)$ . By (3.1) and (3.2), it follows that

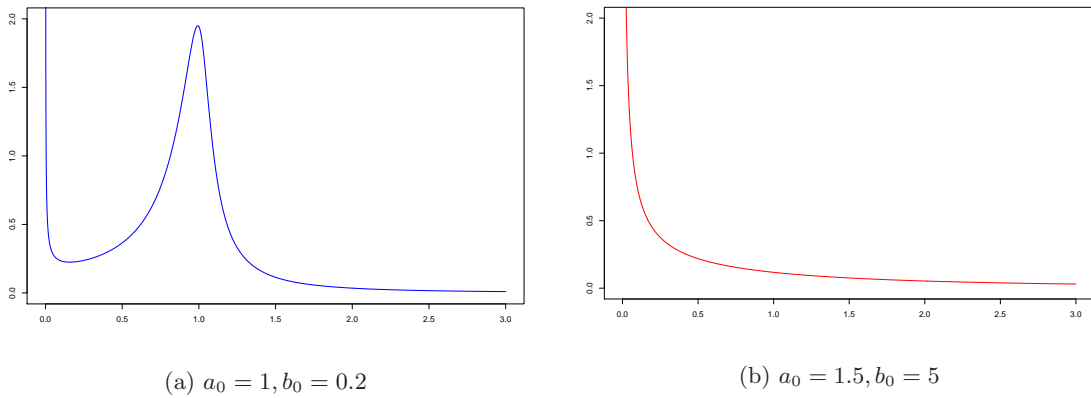


Figure 4.1: Marginal error distribution density functions for different hyperparameters of the gamma prior for the shape parameter  $\theta$ . (a) Marginal error of Leon et al.'s parametric model, and (b) marginal error used in our semiparametric models.

$$\begin{aligned} \mathbb{E}[\log(V)] &= \mathbb{E}[\mathbb{E}[\log(V|\theta)]] = \mathbb{E}\left[\mathbb{E}\left[\frac{W}{\theta} \mid \theta\right]\right] \\ &= \mathbb{E}\left[\frac{1}{\theta} \mathbb{E}[W]\right] = \mathbb{E}\left[-\frac{\gamma}{\theta}\right] = -\gamma \frac{b_0}{a_0 - 1}, \quad a_0 > 1, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[(\log(V))^2] &= \mathbb{E}[\mathbb{E}[(\log(V|\theta))^2]] = \mathbb{E}\left[\mathbb{E}\left[\left(\frac{W}{\theta}\right)^2 \mid \theta\right]\right] = \mathbb{E}\left[\frac{1}{\theta^2} \mathbb{E}[W^2]\right] \\ &= \mathbb{E}\left[\frac{1}{\theta^2} \left(\frac{1}{6}\pi^2 + \gamma^2\right)\right] = \left(\frac{1}{6}\pi^2 + \gamma^2\right) \frac{b_0^2}{(a_0 - 1)(a_0 - 2)}, \quad a_0 > 2, \end{aligned}$$

where  $\gamma = 0.577\dots$  is the Euler constant. Hence we think that their prior on  $\theta$  is too much informative (see the peak around  $v = 1$  in Figure 4.1a), and has too heavy tails since it does not even admit the first log-moment. The marginal error distribution density function with parameters  $a_0 = 1.5$  and  $b_0 = 5$  in Figure 4.1b is less informative and has lighter tails. It can be proved that the asymptote at zero is present for any choices of the



hyperparameters. To avoid the asymptote we could take into account priors with limited support, for instance a uniform prior between  $a$  and  $b$ , where  $a, b > 0$ ; but the drawback is that the posterior is forced to have this limited support.

In Table 4.1 we provide their credibility intervals of the quantiles of the predictive distributions for a new spool, and for Spool 2 and 7. Notice that the intervals for Spool 2 and 7 are quite different, and those of a new random spool are much larger. We guess that this gap between Spool 2 and 7 is due to the microscopic differences between the two spools (see Table 3.1). We fitted the parametric model using the code provided in their article and observed that the results are very strongly affected by Monte Carlo errors; so we will consider only the order of magnitude and not the exact values of their interval estimates.

Table 4.1: Interval estimates of the quantiles of the predictive distributions for the parametric model

Spool	2.5%	50%	97.5%
2	153.2	362.4	732.6
7	56	131	305
new	22.0	671	19290

(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	17.4	28.56	46.42
7	4.72	9.19	17.9
new	1.9	53.7	1479

(b) Median failure time in thousands of hours at 22.5MPa.

## 4.1 Nonparametric error model

We have considered the model (3.7) under the following prior

$$\begin{aligned}
 \beta_0 &\sim N(0, \sigma_0^2 = 0.001) \\
 \beta_1 &\sim N(0, \sigma_1^2 = 0.001) \\
 \alpha_1, \dots, \alpha_J | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\
 \lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2} = 0.1, \frac{\tau_2}{2} = 0.1\right) \\
 \theta_1, \dots, \theta_n | P &\sim P \\
 P &\sim \text{NGG}(\sigma, \omega = 1, \kappa, P_0) \\
 P_0 &\sim \text{Gamma}(a_0 = 1.5, b_0 = 5).
 \end{aligned} \tag{4.2}$$

In this Section we will use those hyperparameters if not otherwise specified, while the NGG processes prior hyperparameters  $\sigma$  and  $\kappa$  will be specified each time. Observe that

the variance of the fixed effects priors are as in the parametric model (4.1), while we take  $\tau := \tau_1 = \tau_2 = 0.2$  to have a priori random effects with less heavy tails.  $P_0$  is the base-line measure of the NGG process prior and we fix it to obtain a marginal error as in in Figure 4.1b.

#### 4.1.1 Interval estimates for a new random spool and for given ones

First we carry out a sensibility analysis with respect to the nonparametric prior hyperparameters  $\sigma$  and  $\kappa$ , providing the credibility intervals of the first percentile of the failure time distribution at stress level 23.4MPa, and of the median failure time when stress is 22.5MPa for a new random spool. We also provide the credibility intervals at the two stress levels for Spool 2 and 7.

In Table 4.1.1, we provide credibility intervals of the quantiles of the predictive distribution for different choices of the NGG process hyperparameters  $\sigma$  and  $\kappa$ . We fixed them in order to have an expected number of clusters of  $\theta_i$ 's between 2 and 8 to reflect our a priori knowledge. We fixed the expected number of cluster equal to 2 because we know that there are at least two groups of spools, those that are close to the nominal characteristics like Spool 2, and those that do not respect the nominal conditions like Spool 7; we chose the other "extreme" 8, since this is exactly the information we have from the data (i.e. the data come exactly from 8 spools). We recall that in this model the number of groups of  $\theta_i$ 's belongs to  $\{1, \dots, n\}$  and does not necessarily represent the number of spools, but the components of the error. Finally we tested the robustness of the model fixing  $\sigma$  and  $\kappa$  such that the expected number of clusters is higher (20.6). We did not use values for  $\sigma$  higher than 0.3 because the number of relevant components of  $P$  increases too much and consequently the computing time explodes. We also noticed numerical problems in the evaluation of the weights of  $P$ .

From Table 4.1.1 it is clear that the median quantile at 22.5MPa is more robust than the 1st percentile at 23.4MPa. We observe a monotone trend as a function of  $\mathcal{I}(\sigma, \kappa)$ . In fact, if  $\mathcal{I}(\sigma, \kappa)$  gets smaller, the credibility intervals moves to the left and gets narrower. As showed in (1.7),  $\mathcal{I}(\sigma, \kappa)$  has a multiplicative effect on the variance of the base-line measure  $P_0$ , and so, if  $\mathcal{I}(\sigma, \kappa)$  is small, we have little variability around  $P_0$ . Moreover, we recall that if  $\mathcal{I}(\sigma, \kappa) \rightarrow 1$ , we obtain the parametric model (3.4), while if  $\mathcal{I}(\sigma, \kappa) \rightarrow 0$ , we obtain the following parametric model

$$\begin{aligned} V_i | \theta_i &\stackrel{\text{iid}}{\sim} Weibull(\theta_i, 1), \quad i = 1, \dots, n, \\ \theta_i &\stackrel{\text{iid}}{\sim} Gamma(a_0, b_0). \end{aligned}$$

This is confirmed by our numerical results; in fact, when  $\mathcal{I}(\sigma, \kappa)$  is close to one, we obtain results comparable to those of Leon et al. (see Table 4.1), while, when  $\mathcal{I}(\sigma, \kappa)$  is close to

Table 4.2: Interval estimates of the quantiles of the predictive distributions for the model with nonparametric error for different hyperparameters of the NGG process prior.

$\sigma$	$\kappa$	$\mathcal{I}(\sigma, \kappa)$	$\mathbb{E}[K_n]$	$\mathbb{E}[K_n \mathbf{T}]$	2.5%	50%	97.5%
0.01	0.2	0.82	2.9	1.3	10.7	557.1	16992.8
0.1	0.1	0.81	2.3	1.2	13.1	583.3	17723.1
0.3	0.03	0.67	4.9	1.2	15.8	609.3	19023.5
0.3	0.3	0.51	7.4	2.0	1.4	329.5	10509.8
0.1	1	0.43	6.9	2.5	0.1	248.1	10220.5
0.01	2	0.33	8.7	3.2	0.0	129.5	6903.5
0.3	3	0.17	20.9	6.6	0.0	0.4	306.0

(a) 1st percentile failure time in hours at 23.4MPa.

$\sigma$	$\kappa$	$\mathcal{I}(\sigma, \kappa)$	$\mathbb{E}[K_n]$	$\mathbb{E}[K_n \mathbf{T}]$	2.5%	50%	97.5%
0.01	0.2	0.82	2.9	1.3	1.9	62.0	1685.7
0.1	0.1	0.81	2.3	1.2	2.2	64.6	1895.3
0.3	0.03	0.67	4.9	1.2	2.5	65.3	1821.8
0.3	0.3	0.51	7.4	2.0	2.0	62.5	1603.4
0.1	1	0.43	6.9	2.5	2.0	64.4	1634.9
0.01	2	0.33	8.7	3.2	2.2	63.8	1797.0
0.3	3	0.17	20.9	6.6	2.5	63.6	1918.6

(b) Median failure time in thousands of hours at 22.5MPa.

zero, the process  $P$  has a very small variance, and so the prior is too strong and the data do not wash out the prior information.

Additionally, we notice that the posterior expected number of clusters of  $\{\theta_i\}_{i=1}^n$  is always smaller than the prior one, and this reduction of number of clusters is more evident for higher values of  $\sigma$ . In fact, as we can see in Figure 4.2, for a given prior expected number of clusters, we have a larger dispersion for higher values of  $\sigma$ .

If the expected number of components  $\mathbb{E}[K_n]$  is much higher than one, the data do not overwhelm the prior information; while if the expected number of components  $\mathbb{E}[K_n]$  is close to one, we obtain similar results to the parametric model of Leon et al.. In short, we think that the nonparametric shape mixture does not improve substantially the predictive capability of the parametric model; therefore a single parametric  $\theta$  as in Leon et al. is enough.

In Table 4.3 we report the interval estimates of the two quantiles of interest for Spool 2 and 7, with  $\sigma = 0.3$  and  $\kappa = 0.3$ . We recall that we are mainly interested in the prediction for a new random spool, but we consider that is interesting as well to compare our results

to those of the parametric model proposed by Leon et al..

Notice that the credibility intervals of the 1st percentile at 23.4MPa are narrower and centered to smaller values respect to the parametric model's ones; while intervals of the median at 22.5MPa are slightly wider and more to the right. We recall that the results of Leon et al. are affected by a large Montecarlo error and hence the differences for the median at 22.MPa are not significant. As we mentioned in Section 3.2.1, Spool 7 does not respect the nominal conditions, while Spool 2 does. We observe a big differences between the two spools in particular at stress level 22.5MPa, and we can guess that the bad behavior of Spool 7 is due to to the fact that it has fewer yarns per fiber than a nominal Kevlar fiber.

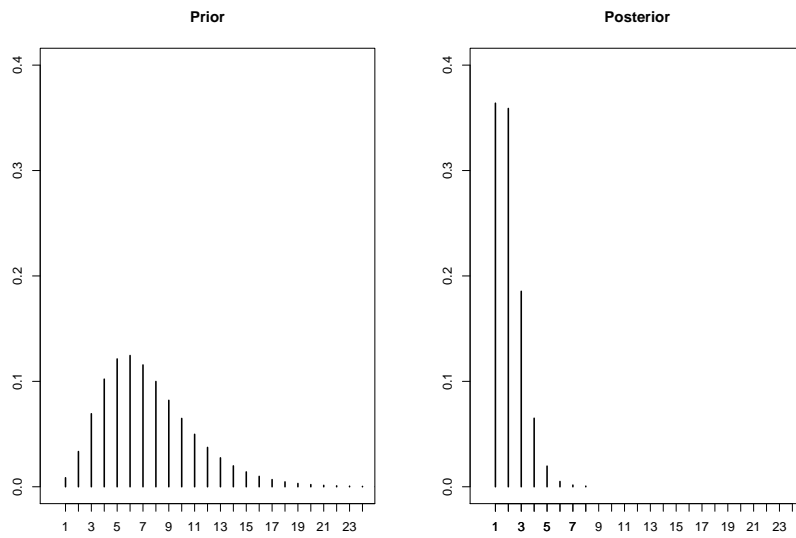
Table 4.3: Interval estimates of the quantiles of the predictive distributions for the model with nonparametric error ( $\sigma = 0.3$ ,  $\kappa = 0.3$ )

Spool	2.5%	50%	97.5%
2	1.9	204.2	530.8
7	0.6	61.9	183.5
new	1.4	329.5	10509.8

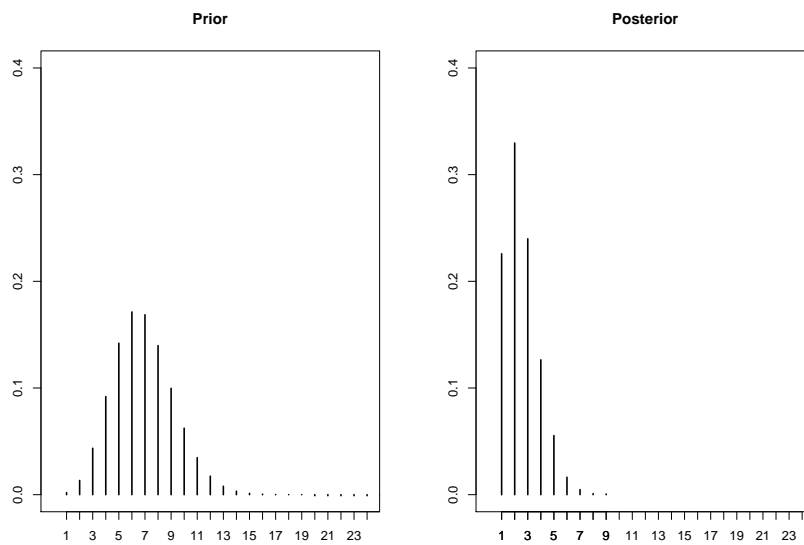
(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	19.2	32.4	59.8
7	4.9	9.8	22.6
new	2.0	62.5	1603.4

(b) Median failure time in thousands of hours at 22.5MPa.



(a)  $\sigma = 0.3, \kappa = 0.3$



(b)  $\sigma = 0.1, \kappa = 1$

Figure 4.2: Prior and posterior number of mixture components for the model with non-parametric error

### 4.1.2 Posterior estimates

In Table 4.4 we provide the posterior estimates of the fixed effects  $\beta_0$  and  $\beta_1$ , the random effects  $\alpha_1, \dots, \alpha_8$ , and the new random effect  $\alpha_{\text{new}}$ . Observe that the stress effect  $\beta_1$  is negative, since it is expected, because high pressure levels reduce the life-time of the Kevlar fiber; and Spool 1,4, and 8 have the best performances, while Spool 3 and 7 the worst ones. The differences between the spools and the new random spool can be seen in the plot of the posterior kernel density (see Figure 4.3). Moreover, all  $\alpha_j$ 's have almost the same standard deviation around 0.60, while the new random effect  $\alpha_{\text{new}}$  has a larger standard deviation (1.58). The high variability of the new random spool is due to the fact that we do not know if it will have good (or even better) performance as Spool 4, mean one, or bad (or even worse) as Spool 7.

Table 4.4: Posterior mean and standard deviation (SD) of the effects for the model with nonparametric error ( $\sigma = 0.3$  and  $\kappa = 0.3$ )

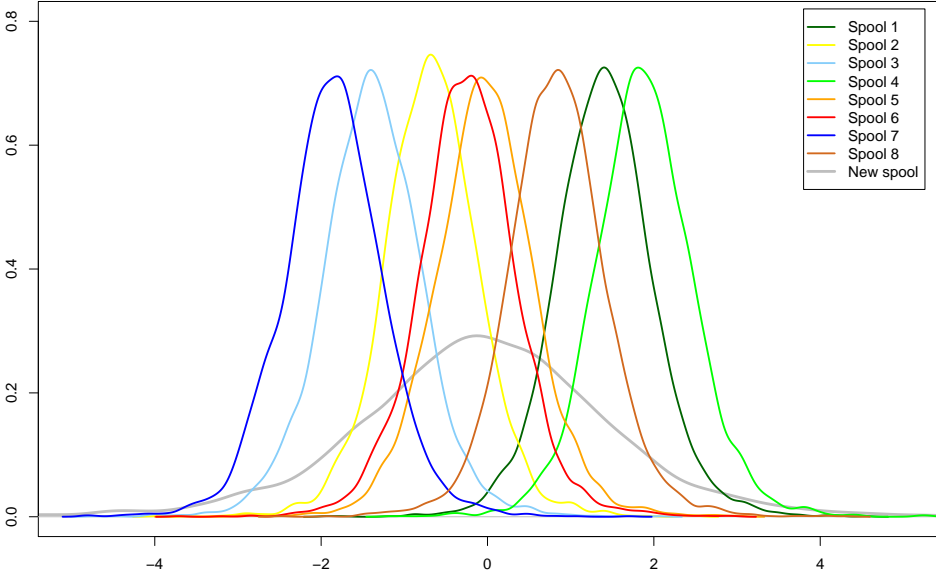
	$\beta_0$	$\beta_1$
Mean	-1.04	-23.26
SD	0.59	1.17

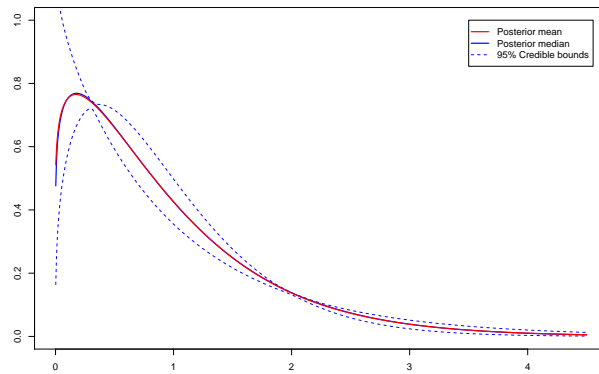
  

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_{\text{new}}$
Mean	1.41	-0.67	-1.38	1.89	-0.03	-0.24	-1.86	0.85	-0.04
SD	0.60	0.59	0.61	0.61	0.62	0.61	0.62	0.60	1.58

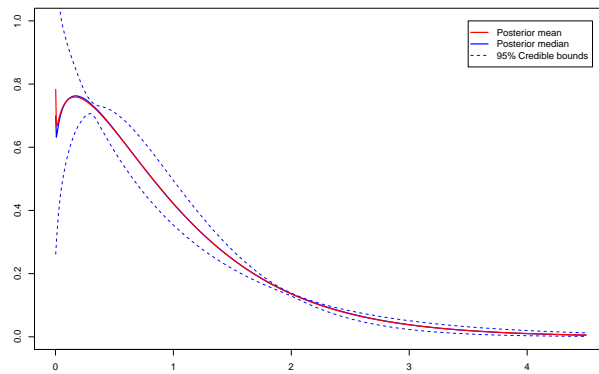
In Figure 4.4, we provide the posterior marginal density distribution function of the error for  $\sigma = 0.3$  and  $\kappa = 0.03, 0.3, 3$ . Notice that for  $\kappa = 3$  the posterior is similar to the base-line measure  $P_0$ ; while in the other two cases, the data moved the mass of the distribution to the right. This phenomenon is more evident for  $\kappa = 0.03$ . For all these three density functions the 95% credible bounds are very wide at zero and particularly narrow around 0.3. This happens because the posterior marginal density function is a mixture of the prior function with asymptote at zero and a density function without asymptote, and those two density distribution functions cross each other around 0.3.

Figure 4.3: Posterior kernel density estimation of the random effects for the model with nonparametric error ( $\sigma = 0.3$  and  $\kappa = 0.3$ )

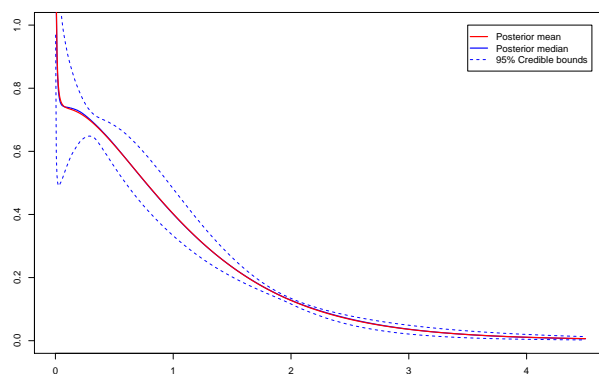




(a)  $\sigma = 0.3$  and  $\kappa = 0.03$



(b)  $\sigma = 0.3$  and  $\kappa = 0.3$



(c)  $\sigma = 0.3$  and  $\kappa = 3$

Figure 4.4: Posterior marginal density distribution functions of the error in the model with nonparametric error



### 4.1.3 Predicted survival functions

In Figure 4.5, we plot the predicted survival functions for a new random spool at stress level 22.5MPa with  $\sigma = 0.3$  and  $\kappa = 0.03, 0.3, 3$ .

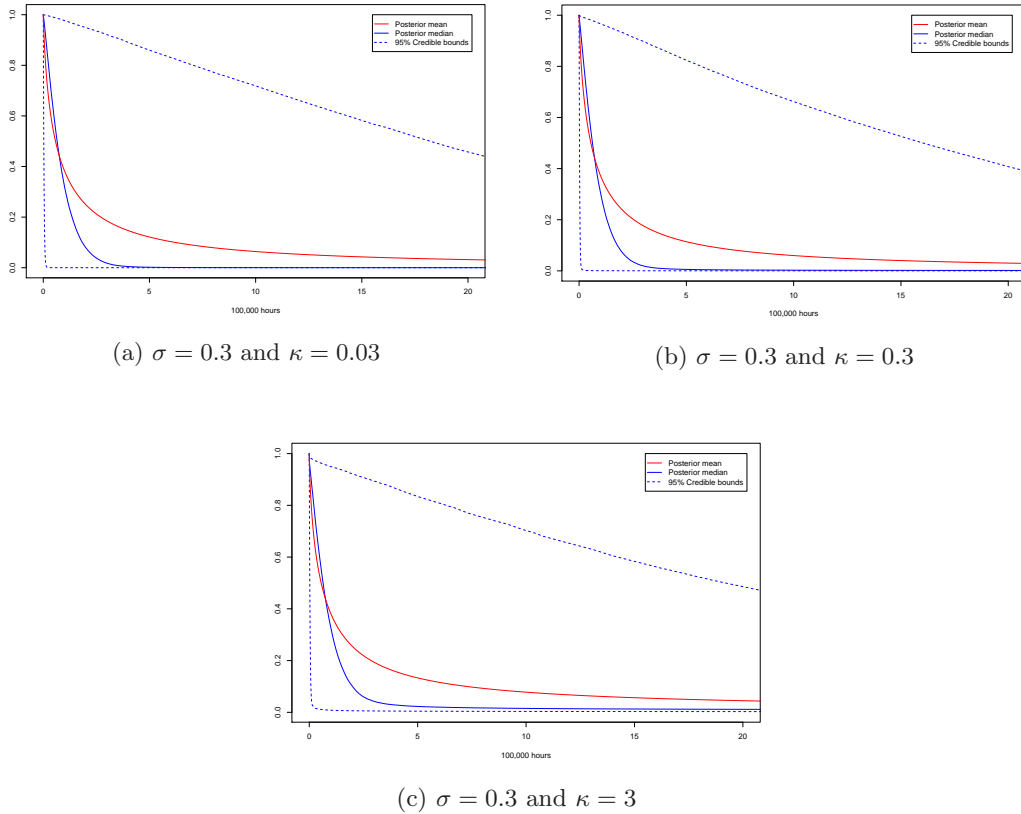


Figure 4.5: Survival functions for a new random spool at 22.5MPa for the model with nonparametric error

In Figure 4.6, we report the predicted survival function for a new random spool at stress level 23.4MPa ( $\sigma = 0.3, \kappa = 0.3$ ), and the Kaplan-Meier estimator. Notice that the Kaplan-Meier point estimator for given quantiles of the failure time is in general higher and its 95% bounds are narrower compared to those of the model with nonparametric error. This big difference in the width of the bounds is due to the fact that the Kaplan-Meier estimator does not treat the spool effects as random, and so it underestimates the variability of a new random spool.

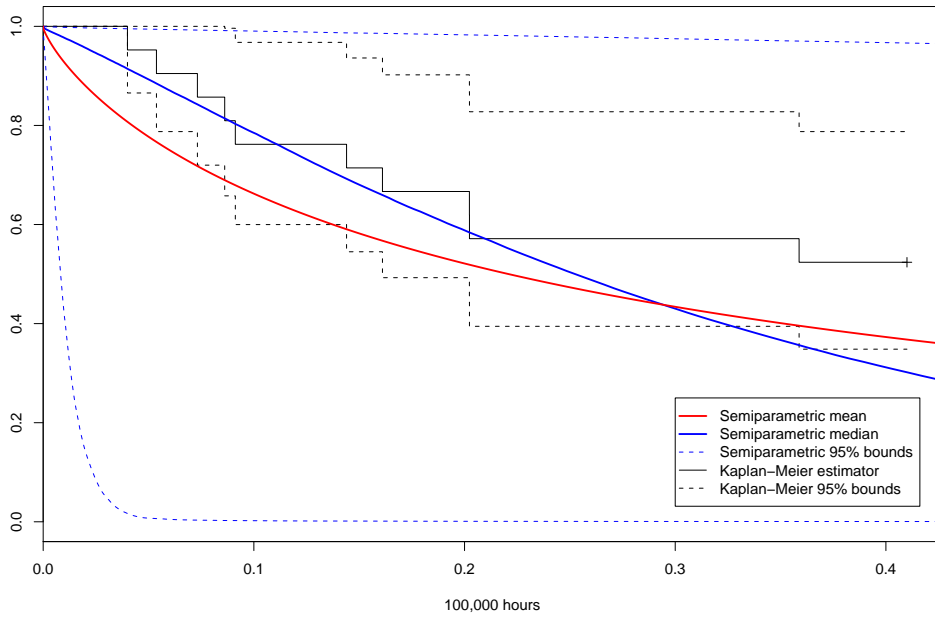


Figure 4.6: Predicted survival functions for a new random spool at 23.4MPa for the model with nonparametric error ( $\sigma = 0.3$ ,  $\kappa = 0.3$ ), and for Kaplan-Meier estimator

#### 4.1.4 Residuals

We computed the Bayesian residuals to evaluate the goodness of fit as in Chaloner [4]. If we write the semiparametric model (3.7) in the additive form as in Section 3.3.1, we have

$$\log T_i = \beta_0 + \beta_1 \log(x_{s,i}) + \alpha_{k[i]} + \frac{W_i}{\theta_i} \quad i = 1, \dots, n, \quad (4.3)$$

where, given the parameters and the covariates,  $W_i$  are a priori independent, identical distributed from the standard Gumbel distribution. The Bayesian residuals for the uncensored failure times are defined as

$$e_i = \mathbb{E}[\theta_i(\log t_i - (\beta_0 + \beta_1 \log(x_{s,i}) + \alpha_{k[i]})) | \mathbf{T}], \quad i = 1, \dots, n - m.$$

Each  $e_i$  can be estimated through MCMC simulation. The following ergodic mean is an estimate of  $e_i$

$$\frac{1}{B} \sum_{b=1}^B \theta_i^{(b)} (\log t_i - (\beta_0^{(b)} + \beta_1^{(b)} \log(x_{s,i}) + \alpha_{k[i]}^{(b)})),$$

where  $\{\theta^{(b)}, \beta_0^{(b)}, \beta_1^{(b)}, \alpha^{(b)}\}_{b=1}^B$  is a sample from the posterior distribution (3.11). In Figure 4.7, we plot the residuals for the uncensored observations with the fitted log failure times on the x-axis. The fitted log failure times are defined as

$$\widehat{\log t_i} = \mathbb{E}[\beta_0 + \beta_1 \log(x_{s,i}) + \alpha_{k[i]} + \frac{W_i}{\theta_i} | \mathbf{T}], \quad i = 1, \dots, n$$

and can be estimated by ergodic means as follow

$$\frac{1}{B} \sum_{b=1}^B (\beta_0^{(b)} + \beta_1^{(b)} \log(x_{s,i}) + \alpha_{k[i]}^{(b)} - \frac{\gamma}{\theta_i^{(b)}}), \quad i = 1, \dots, n,$$

where  $\gamma = 0.577\dots$  is the Euler constant. The dashed lines indicate the 95% bounds of the standard Gumbel distribution, and 95% interval estimates is provided for each residual  $e_i$ . Observe that all but one residuals are in the 95% region of the standard Gumbel.

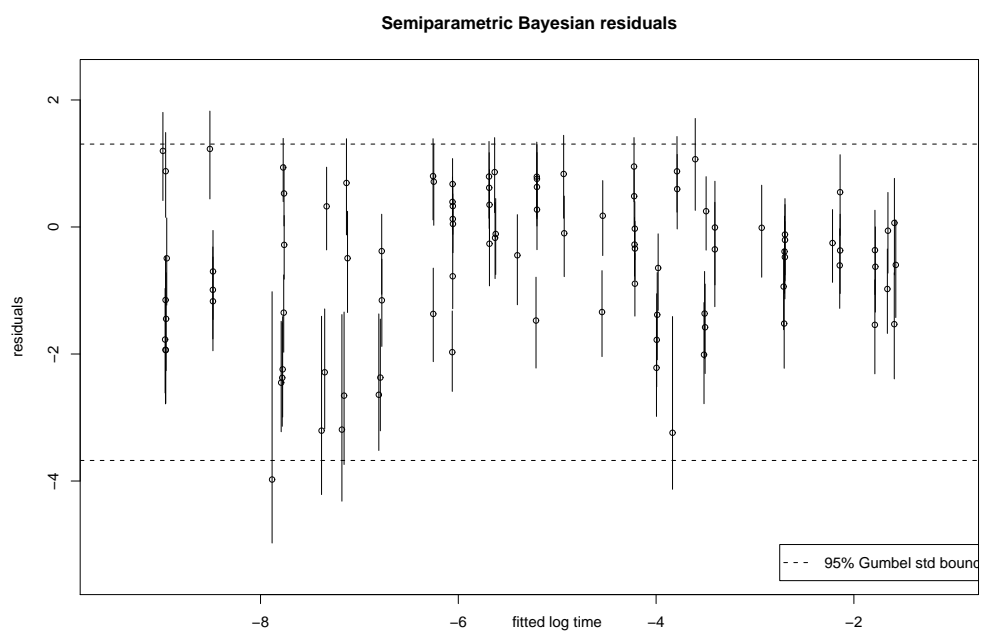
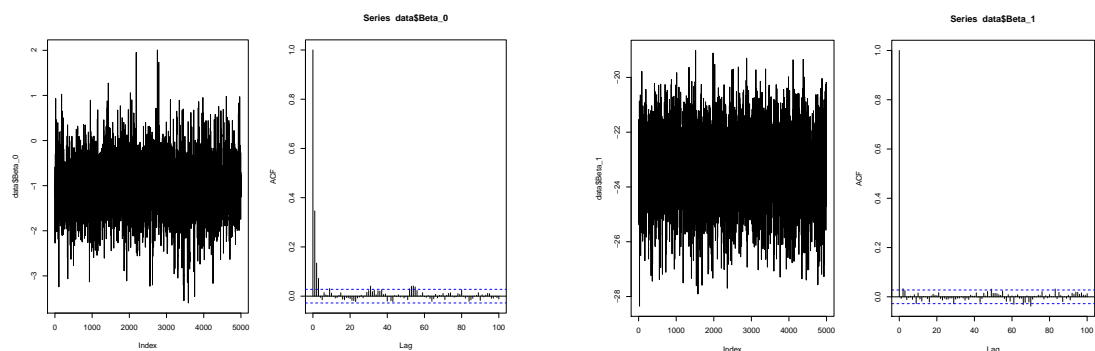


Figure 4.7: Bayesian residuals for the uncensored failure times under the model with nonparametric error ( $\sigma = 0.3$ ,  $\kappa = 0.3$ ).

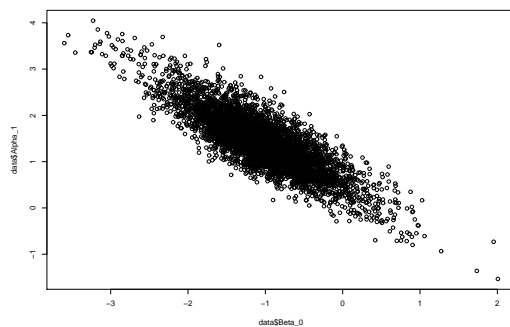
### 4.1.5 Computation details

Posterior and predictive estimates are computed via the Gibbs sampler algorithm presented in Section 3.3.2. We run the algorithm for 350,000 iterations, while the first 100,000 iterations were discarded and we used a thinning of 50 to reduce the autocorrelation of the Markov chain. The final sample size was 5,000. We run longer chains but we do not obtain any relevant reduction of the Monte-Carlo error, and some diagnostic convergence tests were done. In Figures 4.8a and 4.8b, we report traces and estimated autocorrelation functions of  $\beta_0$  and  $\beta_1$ , respectively.



(a) Traces and estimated autocorrelation functions of  $\beta_0$

(b) Traces and estimated autocorrelation functions of  $\beta_1$



(c) Scatterplot of  $\beta_0$  and  $\alpha_1$

Figure 4.8: Markov chain sample of  $\beta_0$ ,  $\beta_1$  and  $\alpha_1$

As already mentioned in Section 3.3.3, we notice a strong correlation between  $\beta_0$  and each  $\alpha_j$  (see Figure 4.8c representing the scatterplot of  $\beta_0$  and  $\alpha_1$ ).

To estimate the cumulative distribution function of the survival time, first we sample the cumulative distribution function of the survival time on a grid and then we make a linear interpolation to fit the curve. Once we have the fitted curve, we estimate any given quantiles of the failure time distribution or plot the survival function. The local error of this approximation depends on the distance between two consecutive sampling points

and the second derivative of the approximated function. We observed that the cumulative distribution function is quite irregular around zero and so we used an anisotropic grid with more points around zero to control the error. In particular, for the stress level 23.4MPa we make a grid with 800 points between 0 and 5 milion hours, imposing an exponentially increasing distance between the points; while for the stress level 22.5MPa we make a similar grid with extremes 0 and 10 milion hours. For example, in Table 4.5 we report the interval estimate obtained for grids with different number of sampling points, with  $\sigma = 0.3$  and  $\kappa = 0.1$ . Notice that the results for 800 and 8000 are similar.

Table 4.5: Interval estimates of the quantiles of the predictive distributions for different number of grid points for the model with nonparametric error ( $\sigma = 0.3$ ,  $\kappa = 0.1$ ).

Points	2.5%	50%	97.5%
80	15.8	609.1	19021.5
800	15.8	609.3	19023.5
8000	15.8	609.3	19023.6

(a) 1st percentile failure time in hours at 23.4MPa.

Points	2.5%	50%	97.5%
80	2.5	65.4	1823.2
800	2.5	65.3	1821.8
8000	2.5	65.3	1821.8

(b) Median failure time in thousands of hours at 22.5MPa.

## 4.2 Nonparametric random effects model

In this section we present the results obtained for the model with nonparametric random effects (3.14)- (3.15). We will use the following hyperparameters if not indicated

$$\begin{aligned}
\theta &\sim \text{Gamma}(a_0 = 1.5, b_0 = 5) \\
\beta_1 &\sim N(0, \sigma_1^2 = 1000) \\
\alpha_1, \dots, \alpha_J | P &\stackrel{\text{iid}}{\sim} P \\
P &\sim \text{NGG}(\sigma, \omega = 1, \kappa, P_0) \\
P_0 | \mu, \lambda &\sim N(\mu, \lambda) \\
\lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2} = 0.1, \frac{\tau_2}{2} = 0.1\right) \\
\mu &\sim N(0, \sigma_0^2 = 1000),
\end{aligned} \tag{4.4}$$

while the process prior hyperparameters  $\sigma$  and  $\kappa$  will be specified each time. We recall that the mean effect  $\beta_0$  was eliminated in the parameterization, but a prior on the mean  $\mu$  of the base-line measure  $P_0$  of the random effects was introduced. Since  $\mu$  has the same role of  $\beta_0$ , here  $\mu$  has the same hyperparameters of  $\beta_0$  in the model with nonparametric error (4.2). The priors on  $\theta$  and on  $\lambda$  are as in (4.2) (see Section 4.1 for an explanation about the hyperparameters choices).

### 4.2.1 Interval estimates for a new random spool and for given ones

First, we make a sensitivity analysis with respect to the nonparametric prior's hyperparameters  $\sigma$  and  $\kappa$ . We provide the 95% intervals estimates of the first percentile of the failure time distribution at stress level 23.4MPa, and of the median when stress is 22.5MPa for a new random spool. Then we give credibility intervals for Spool 2 and 7.

In Table 4.7, we provide the credibility intervals of the quantiles of the predictive distribution for different choices of the hyperparameters of the process  $\sigma$  and  $\kappa$ . We recall that  $K_J$ , the number of clusters of  $\alpha_j$ 's, is random variable with support  $\{1, \dots, J = 8\}$ . We chose  $\sigma$  and  $\kappa$  in order to have an expected number of clusters around 1, 2 and 4. We take 2 because we know that there are at least two groups of spools, those that are close to the nominal characteristics like Spool 2, and those that do not respect the nominal conditions like Spool 7; then we take 1 and 4 to test the robustness of the prior. Finally we fixed  $\sigma = 0.1$  and  $\kappa = 100$  to get close to the parametric model of Leon et al.. In fact, with fixed  $\sigma$  and  $\kappa \rightarrow +\infty$ , we obtain the following parametric model

$$\begin{aligned}
\alpha_1, \dots, \alpha_J | \lambda &\stackrel{\text{iid}}{\sim} N(0, \lambda) \\
\lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2} = 0.1, \frac{\tau_2}{2} = 0.1\right).
\end{aligned} \tag{4.5}$$

Both quantiles are sensitive to the values of the hyperparameters: in particular the intervals get larger and the medians get higher under a larger a priori expected number of clusters. The intervals obtained are narrower than those obtained by the parametric model formula and by the semiparametric model with nonparametric error (3.14)- (3.15). As we expected, for  $\sigma = 0.1$  and  $\kappa = 100$ , we obtain results similar to those of Leon et al.. Moreover, we observe that in all cases the posterior mean number of clusters is bigger than the prior one, and this effect is more relevant with high values of  $\sigma$ .

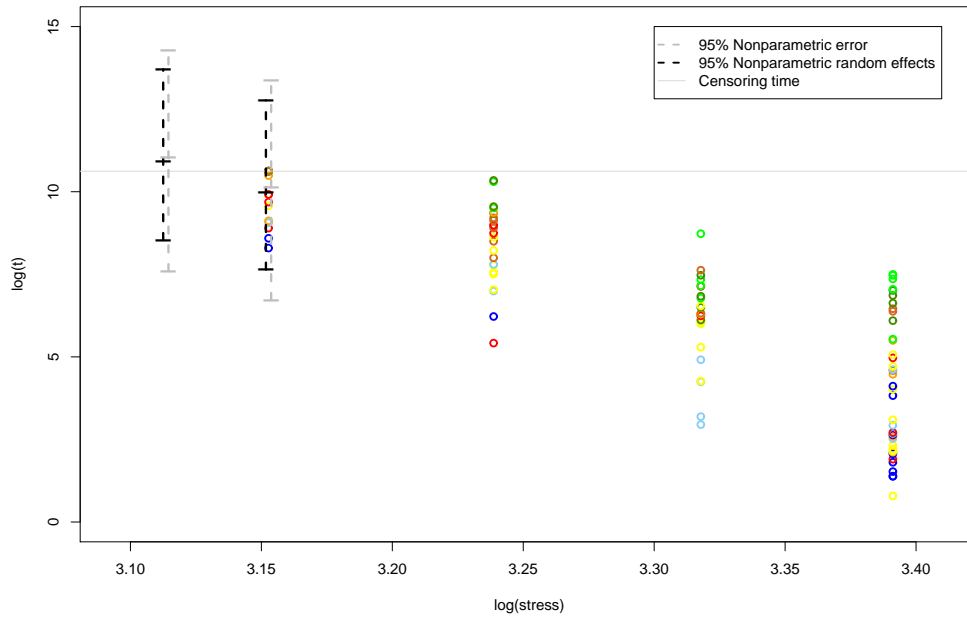


Figure 4.9: Scatterplot of the log-survival times against the log-stress for the two semi-parametric models.

In Figure 4.9 a scatterplot of the log-survival-time against the log-stress is shown together with the estimates of the median survival time at different log-stress levels. In particular, at log-stress equal to  $\log(22.5)=3.11$  and  $\log(23.4)=3.15$ , we provide the credible intervals of the predictive survival median time. In the dashed black line corresponds to the estimates under the model with nonparametric random effects ( $\sigma = 0.3$ ,  $\kappa = 1.2$ ), and in the dashed grey line corresponds to the model with nonparametric error ( $\sigma = 0.3$ ,  $\kappa = 0.3$ ). The horizontal line identifies the censoring time, and the colored points are the failure times, where each color is associated to a given spool as in Figure 4.4. In the log-scale we notice the additive effect of the log-stress on the failure time. Moreover, since each spool has a different color we notice the differences between them: for instance Spool 3 (bluesky) and 7 (blue) are generally the lowest dots, while Spool 1 (dark green)



and 4 (brilliant green) are the highest. Finally, observe that the credible intervals of the predictive survival median time at both stress levels for the model with nonparametric error are larger than those of the model with nonparametric random effects.

In Table 4.6 we report the interval estimates of the two quantiles of interest for Spool 2 and 7, where we fixed  $\sigma = 0.3$  and  $\kappa = 1.2$ . The results for Spool 2 and 7 are similar to those of Leon et al. (see Table 4.1), while both the interval estimates for a new random spool are narrower and with higher median.

Table 4.6: Interval estimates of the quantiles of the predictive distributions for the model with nonparametric random effects ( $\sigma = 0.3$ ,  $\kappa = 1.2$ )

Spool	2.5%	50%	97.5%
2	130.6	364.4	818.8
7	40.6	116.9	289.6
new	41.0	572.6	10027.7

(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	19.2	36.0	64.5
7	5.5	11.5	24.8
new	5.1	55.4	901.3

(b) Median failure time in thousands of hours at 22.5MPa.

Table 4.7: Interval estimates of the quantiles of the predictive distributions for the model with nonparametric random effects for different hyperparameters of the NGG process prior.

$\sigma$	$\kappa$	$\mathcal{I}(\sigma, \kappa)$	$\mathbb{E}[K_J]$	$\mathbb{E}[K_J \mathbf{T}]$	2.5%	50%	97.5%
0.01	0.01	0.98	1.0	2.8	57.48	358.8	4330.9
0.01	0.5	0.66	2.0	4.0	49.1	473.7	5867.0
0.1	0.3	0.68	2.0	4.1	53.7	465.9	6620.2
0.3	0.09	0.62	2.0	5.0	45.2	524.4	7285.3
0.5	0.1	0.43	3.5	6.1	42.8	556.1	9625.8
0.01	2.5	0.28	4.0	5.4	44.9	576.7	8430.6
0.1	2	0.29	4.0	5.4	46.1	556.5	8506.7
0.3	1.2	0.30	4.1	5.7	41.0	572.6	10027.7
0.5	0.3	0.32	4.1	6.2	42.3	587.4	10314.9
0.7	0.01	0.29	4.7	7.0	33.2	642.4	13593.7
0.1	100	0.009	7.8	7.8	23.5	676.7	18575.1

(a) 1st percentile failure time in hours at 23.4MPa.

$\sigma$	$\kappa$	$\mathcal{I}(\sigma, \kappa)$	$\mathbb{E}[K_J]$	$\mathbb{E}[K_J \mathbf{T}]$	2.5%	50%	97.5%
0.01	0.01	0.98	1.1	2.8	8.6	44.5	505.1
0.01	0.5	0.66	2.0	4.0	7.2	49.3	576.1
0.1	0.3	0.68	2.0	4.1	7.7	48.3	644.2
0.3	0.09	0.62	2.0	5.0	6.5	51.2	673.5
0.5	0.1	0.43	3.5	6.1	4.9	53.2	906.8
0.01	2.5	0.28	4.0	5.4	5.5	54.2	763.3
0.1	2	0.29	4.0	5.4	5.9	53.3	790.3
0.3	1.2	0.30	4.1	5.7	5.1	55.4	901.3
0.5	0.3	0.32	4.1	6.2	5.0	55.2	905.1
0.7	0.01	0.29	4.7	7.0	3.6	59.6	1202.0
0.1	100	0.009	7.8	7.8	2.3	61.4	1708.3

(b) Median failure time in thousands of hours at 22.5MPa.

## 4.2.2 Posterior estimates

In Table 4.8 we provide the posterior estimates of the fixed effects  $\beta_1$ , the random effects  $\alpha_1, \dots, \alpha_8$ , and the new random effect  $\alpha_{\text{new}}$ . Notice that all  $\alpha_j$ s have almost the same standard deviation around 0.3, while the new random effect  $\alpha_{\text{new}}$  has a larger standard deviation (1.68). Moreover, we can compare these posterior estimates to those of the model with nonparametric error of Table 4.4. Observe that the log-stress effect  $\beta_1$  has similar mean and SD of that of the model with nonparametric error. Even if the random effects in the two models have been modeled differently, the standard deviations of the random effects are smaller here than those in Table 4.4 corresponding to the model with nonparametric error. This shrinkage effect may be a consequence of the increased flexibility due to the nonparametric random effects.

Table 4.8: Posterior mean and standard deviation (SD) of the effects for the model with nonparametric random effects ( $\sigma = 0.3$  and  $\kappa = 1.2$ )

	$\beta_1$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_{\text{new}}$
Mean	-23.42	0.46	-1.62	-2.47	0.77	-1.18	-1.34	-2.75	-0.05	-1.01
SD	1.17	0.35	0.28	0.37	0.36	0.36	0.29	0.36	0.37	1.68

In Figure 4.10 we plot the kernel density estimation of the posterior distribution of the random-effects parameters. Roughly, we identify 3 groups of spools and consequently the posterior density estimation of the new random spool has 3 local maxima. Notice that the posterior expected number of clusters is higher than 3, but we recall that often in Bayesian mixture models several components make each cluster.

Figure 4.11 displays the kernel density estimation of the posterior of the shape parameter  $\theta$  and the marginal posterior density distribution function of the error for  $\sigma = 0.3$  and  $\kappa = 1.2$ . Figure 4.11a shows that the data overwhelm the prior information on  $\theta$ . The 95% interval estimates for the shape parameter  $\theta$  is  $[0.95, 1.32]$  and so the failure rate is constant or slightly increasing with time. As we noticed in the model with nonparametric error (see Figure 4.4), the interval estimates of the posterior marginal density distribution of the error are particularly large at zero, and narrow around 0.3.

Figure 4.10: Posterior kernel density estimation of the random effects for the model with nonparametric random effects ( $\sigma = 0.3$  and  $\kappa = 1.2$ )

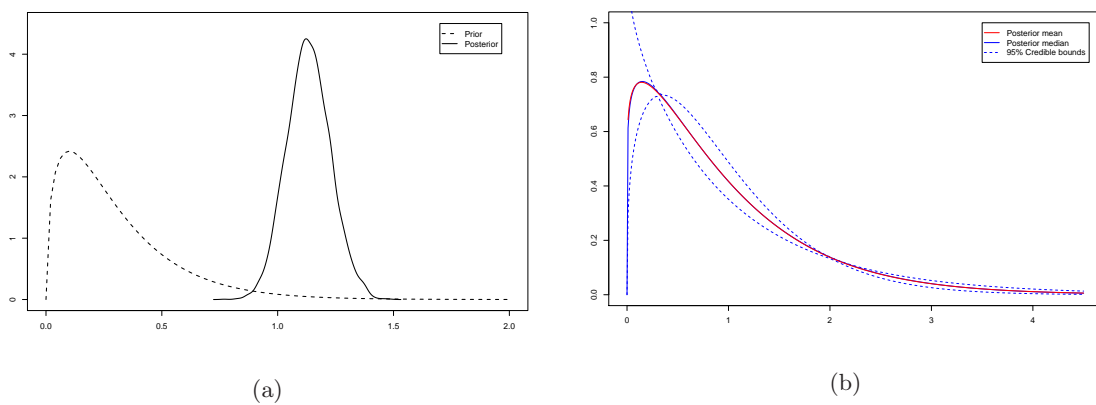
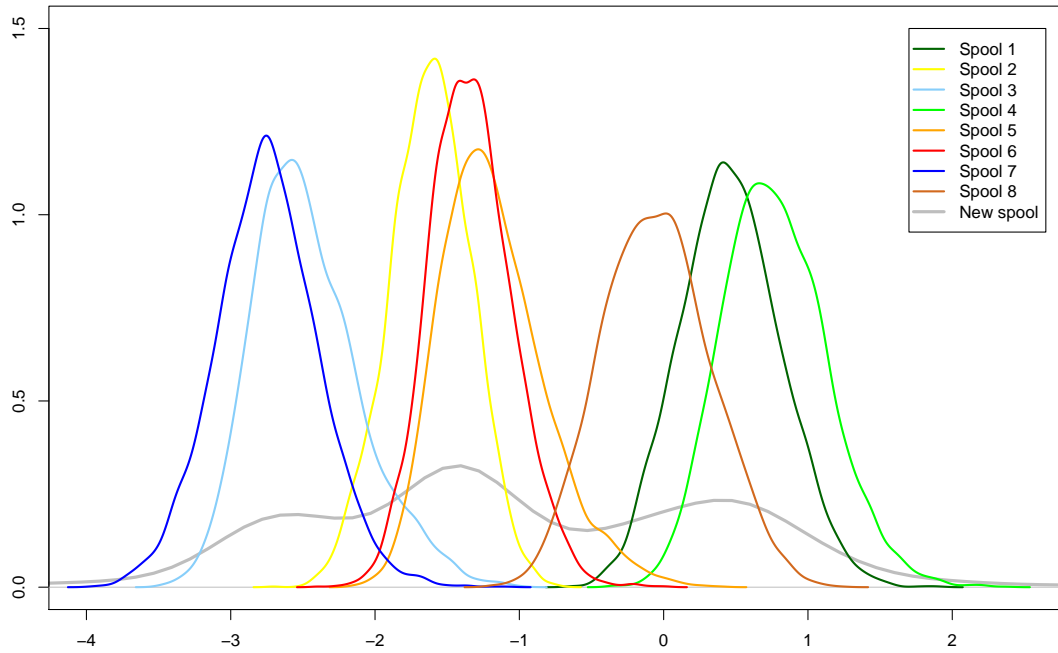
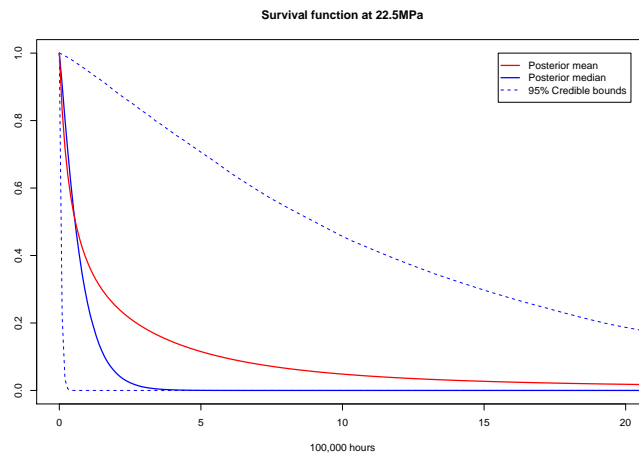


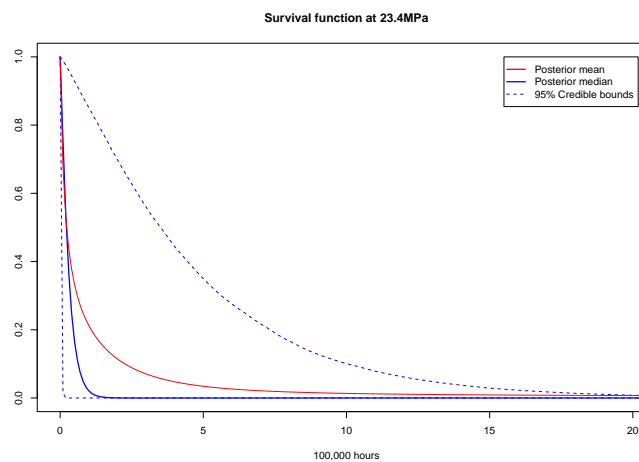
Figure 4.11: (a) Posterior kernel density estimation of the shape parameter  $\theta$  and (b) posterior marginal density distribution function of the error for the model with nonparametric error, where  $\sigma = 0.3$  and  $\kappa = 1.2$ .

### 4.2.3 Predicted survival functions

In Figure 4.2.3, we plot the predicted survival functions for a new random spool at stress level 22.5MPa and 23.4MPa with  $\sigma = 0.3$  and  $\kappa = 1.2$ . Notice that the 2.5% quantile of the survival distribution at stress level 23.4MPa decreases faster than those at stress level 22.5MPa.



(a) Stress level 22.5MPa



(b) Stress level 23.4MPa

Figure 4.12: Predicted survival functions for a new random spool for the model with nonparametric random effects ( $\sigma = 0.3$  and  $\kappa = 1.2$ )

#### 4.2.4 Residuals

We computed the Bayesian residuals to evaluate the goodness-of-fit of (3.14)-(3.15) as we did in Sections 4.1.4. In Figure 4.13 we plot the residuals of uncensored times against the fitted log failure times, and we do not notice relevant differences between the two semi-parametric models (compare to Figure 4.7). The dashed lines indicate the 95% credible intervals of the standard Gumbel distribution, and for each residual  $e_i$  a 95% interval estimates is provided. We observe that only 2 of the 97 residuals are outside the 95% region of the standard Gumbel. We do not see any obvious pattern in the residuals.

Figure 4.14 displays the residuals against the expected quantiles of the standard Gumbel at all stress levels. Notice that the residuals' distribution at stress level 25.5MPa is skewed to the left, while the residuals at stress level 23.4MPa have lighter tails than expected, and so a posteriori the Gumbel assumption is violated.

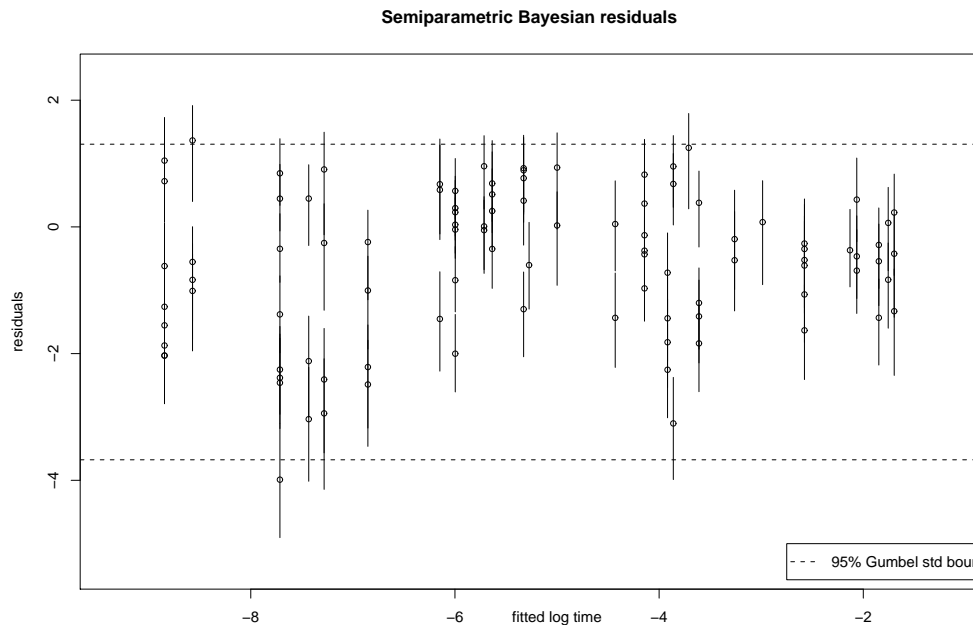


Figure 4.13: Bayesian residuals against the fitted log failure times

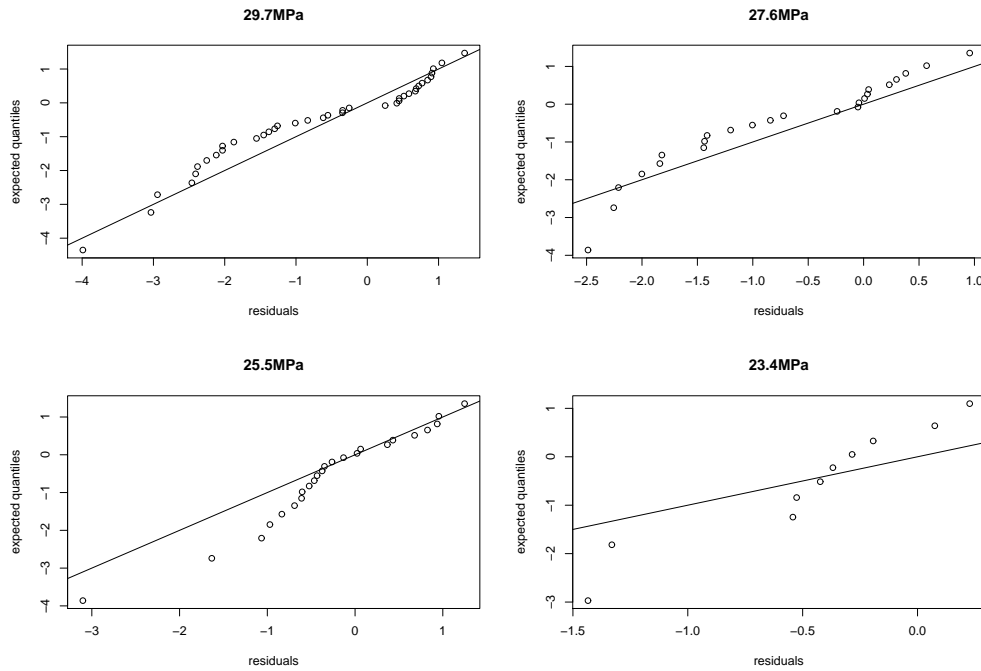


Figure 4.14: Bayesian residuals against the quantiles of the Gumbel standard at the four stress levels

### 4.2.5 Computation details

Posterior and predictive estimates are computed by the MCMC scheme presented in Section 3.3.3. We run the algorithm for 57,000 iterations, with a burn-in of 7,000 iterations and a thinning of 10 to reduce the autocorrelation of the Markov chain. The final sample of size was 5,000. We run longer chains without obtaining any relevant reduction of the Monte-Carlo error, and some diagnostic convergence tests were done. Notice that the Markov chain is less autocorrelated than the one corresponding to the model with non-parametric error (see Figure 4.8c); this is the reason why we reduced the thinning to 10. In Figures 4.15a and 4.15b, we report traces plots and autocorrelation functions of  $\beta_1$  and  $\alpha_1$ , respectively.

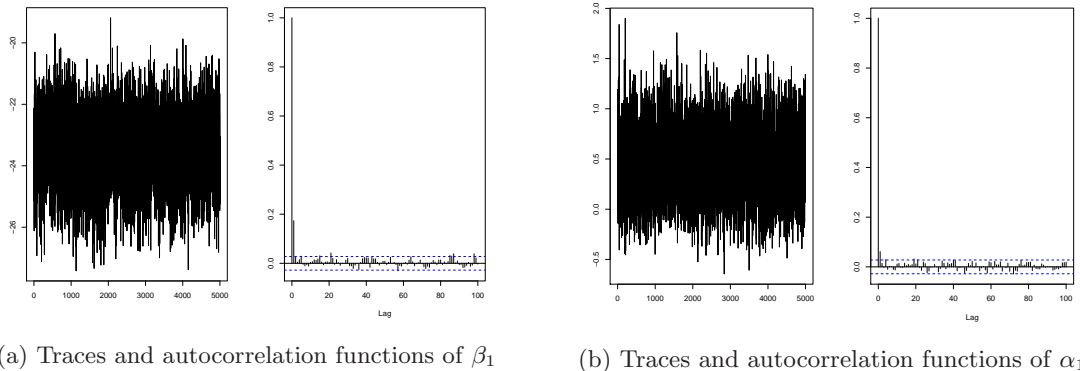


Figure 4.15: Markov chain sample of  $\beta_1$  and  $\alpha_1$

### 4.3 DPpackage: gamma-distributed error model

Here we present the results obtained using *DPpackage* for the model with gamma-distributed error 3.19. We used the following hyperparameters

$$\begin{aligned}
\theta &\sim \text{Gamma}(a_0, b_0) \\
\beta_1 &\sim N(0, \sigma_1^2 = 1000) \\
\alpha_1, \dots, \alpha_J | P &\stackrel{\text{iid}}{\sim} P \\
P &\sim DP(aP_0) \\
a &\sim \text{Gamma}(a_1, b_1) \\
P_0 | \mu, \lambda &\sim N(\mu, \lambda) \\
\lambda &\sim \text{InvGamma}\left(\frac{\tau_1}{2} = 0.1, \frac{\tau_2}{2} = 0.1\right) \\
\mu &\sim N(0, \sigma_0^2 = 1000).
\end{aligned} \tag{4.6}$$

The prior for  $\beta_1$ ,  $\lambda$  and  $\mu$  are as in (4.4); while for the priors  $\theta$  and  $a$  we will make a sensitivity analysis.

We do not have analytical expression of density distribution function of the marginal error  $V$ , but we know its mean and variance,

$$\begin{aligned}
\mathbb{E}[V] &= \mathbb{E}[\mathbb{E}[V|\theta]] = \mathbb{E}[1] = 1 \\
\text{Var}[V] &= \mathbb{E}[\text{Var}[V|\theta]] + \text{Var}[\mathbb{E}[V|\theta]] = \mathbb{E}[\theta] + \text{Var}[1] = \frac{a_0}{b_0}.
\end{aligned} \tag{4.7}$$

Additionally, we can evaluate its density distribution function by numerical integration. In Figure 4.3, we plot four different density distribution functions of the marginal error. The two functions on the right are similar to the marginal error that we used for the two semiparametric models (see Figure 4.1b), while those on the left have a pick as the marginal error in the parametric model proposed by Leon et. al. (see Figure 4.1a).



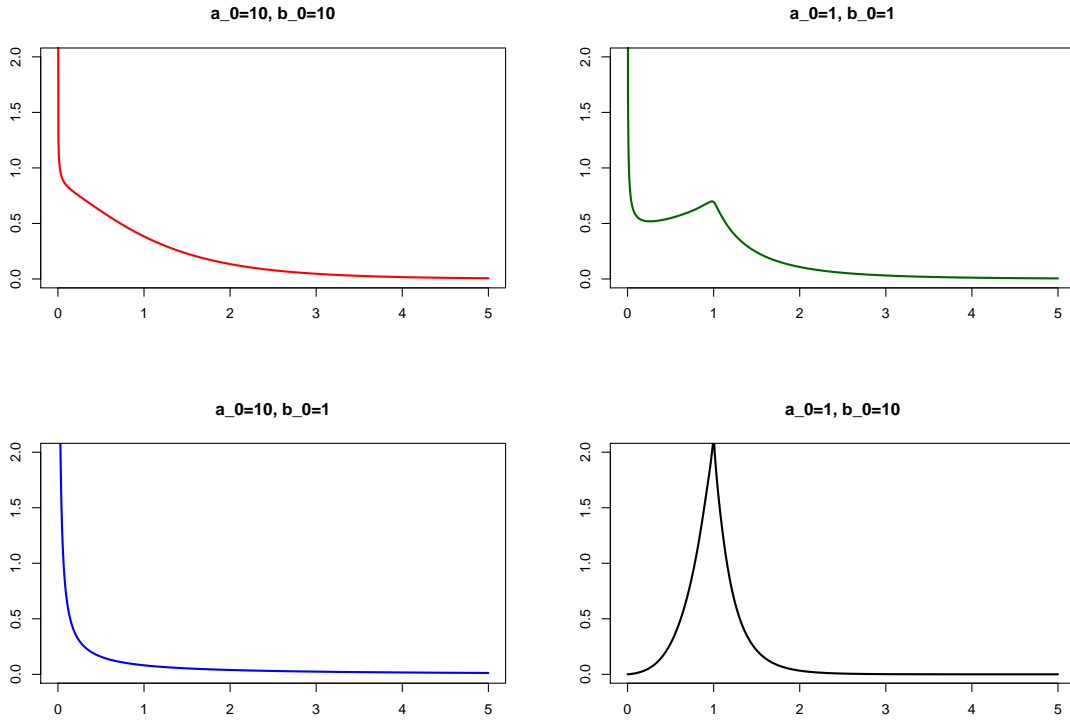


Figure 4.16: Density distribution function of the marginal error  $V$  for different choices of the hyperparameters  $a_0$  and  $b_0$

### 4.3.1 Interval estimates for a new random spool and for given ones

First we give the 95% interval estimates of the two quantiles of interest for a new random spool, for different choices of the hyperparameters, and then for Spool 2 and 7.

In Table 4.9, we fix  $a \sim Gamma(a_1 = 1, b_1 = 1)$  and we vary the hyperparameters of the dispersion parameter  $\theta$  as in Figure 4.3. Notice that the results obtained are comparable with those of the model with nonparametric random effects (see Table 4.7). The interval estimates of median at 22.5MPa are more robust than the 1st percentile at 23.4MPa, and the results for  $a_0 = 1$  and  $b_0 = 10$  are effected by the strong prior information.

In Figure 4.10 we report the results of the robustness analysis on the mass parameter  $a$ , where we fix  $\theta \sim Gamma(a_0 = 1, b_0 = 1)$ . Observe that both the median of the two quantiles are robust to the choice of the hyperparameters.

Table 4.9: Interval estimates of the quantiles of the predictive distributions for the model with gamma error for different hyperparameters of the dispersion parameter  $\theta$  ( $a \sim \text{Gamma}(a_1 = 1, b_1 = 1)$ ).

$a_0$	$b_0$	$\mathbb{E}[\theta]$	$\text{Var}[\theta]$	2.5%	50%	97.5%
10	1	10	10	74.4	885.6	13005.4
1	1	1	1	43.2	583.0	8316.0
10	10	1	0.1	42.0	525.2	8130.3
1	10	1	0.01	21.7	295.6	4908.3

(a) 1st percentile failure time in hours at 23.4MPa.

$a_0$	$b_0$	$\mathbb{E}[\theta]$	$\text{Var}[\theta]$	2.5%	50%	97.5%
10	1	10	10	5.2	47.8	675.7
1	1	1	1	5.0	47.7	601.5
10	10	1	0.1	5.1	47.2	688.7
1	10	1	0.01	4.5	43.6	587.2

(b) Median failure time in thousands of hours at 22.5MPa.

Table 4.10: Interval estimates of the quantiles of the predictive distributions for the model with gamma error for different hyperparameters of the mass parameter  $a$  ( $\theta \sim \text{Gamma}(a_0 = 1, b_0 = 1)$ ).

$a_1$	$b_1$	$\mathbb{E}[a]$	$\text{Var}[a]$	2.5%	50%	97.5%
10	1	10	10	27.2	629.3	13593.5
1	1	1	1	43.2	583.0	8316.0
10	10	1	0.1	43.9	530.8	7498.9

(a) 1st percentile failure time in hours at 23.4MPa.

$a_1$	$b_1$	$\mathbb{E}[a]$	$\text{Var}[a]$	2.5%	50%	97.5%
10	1	10	10	2.7	49.8	1048.2
1	1	1	1	5.0	47.7	601.5
10	10	1	0.1	5.7	44.0	578.3

(b) Median failure time in thousands of hours at 22.5MPa.

In Table 4.11 we provide the same credible intervals for a new random spool, and for Spool 2 and 7; where we fixed the dispersion parameter  $\theta \sim \text{Gamma}(a_0 = 2, b_0 = 2)$  and the mass parameter  $a \sim \text{Gamma}(a_1 = 1, b_1 = 1)$ . Notice that the results obtained are

comparable to those of the model with nonparametric random effect (see Table 4.6).

Table 4.11: Interval estimates of the quantiles of the predictive distributions for the model with gamma error ( $\theta \sim \text{Gamma}(a_0 = 1, b_0 = 1)$  and  $a \sim \text{Gamma}(a_1 = 1, b_1 = 1)$ )

Spool	2.5%	50%	97.5%
2	113.9	366.2	936.7
7	34.1	113.4	300.5
new	43.2	583.0	8316.0

(a) 1st percentile failure time in hours at 23.4MPa.

Spool	2.5%	50%	97.5%
2	16.6	32.6	58.6
7	4.5	9.9	20.7
new	5.0	47.7	601.5

(b) Median failure time in thousands of hours at 22.5MPa.

### 4.3.2 Posterior estimates

In Table 4.12 we provide the posterior estimates of the fixed effect  $\beta_1$ , the random effects  $\alpha_1, \dots, \alpha_J$  and the new random effect  $\alpha_{\text{new}}$ . The random effects has standard deviation around 0.35, while the new random effects has larger variability (1.53). Notice that we obtain estimates similar to those of the model with nonparametric random effects (see Table 4.8).

Table 4.12: Posterior mean and standard deviation (SD) of the effects for the model with gamma error ( $\theta \sim \text{Gamma}(a_0 = 1, b_0 = 1)$  and  $a \sim \text{Gamma}(a_1 = 1, b_1 = 1)$ )

	$\beta_1$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_{\text{new}}$
Mean	-22.76	0.25	-1.74	-2.65	0.52	-1.32	-1.47	-2.91	-0.29	-1.19
SD	1.23	0.37	0.29	0.38	0.37	0.38	0.30	0.35	0.42	1.53

### 4.3.3 Computation details

*DPglm* computes a sample from the posterior distribution through a Polya-urn scheme algorithm. In input we provide the data (the response variables and the design matrix), some parameters on the MCMC (the initial state, the number of iterations and the thinning), the hyperparameters of the priors and the model (the fixed and the random effects, the error distribution and the link-function). Here, it is a script:

```
### DPPackage: DPglm ###
library(DPpackage)
### DATA
data <- data.frame(times=times, stress=stress, spool=spool)
### INITIAL STATE
state <- NULL
### MCMC PARAMETERS
mcmc <- list(nburn=nburn, nsave=nsave, nskip=nskip, ndisplay=ndisplay)
### PRIOR INFORMATION
prior <- list(a0=a0, b0=b0, nu0=nu0, tinv=tinv, tau1=tau1, tau2=tau2,
             mub=mub, Sb=Sb, beta0=beta0, Sbeta0=Sbeta0)
### FIT THE MODEL
fit <- DPglm(fixed=times ~ stress, random=~1|spool, family=Gamma(log),
            prior=prior, mcmc=mcmc, state=state, status=TRUE, data=data)
```

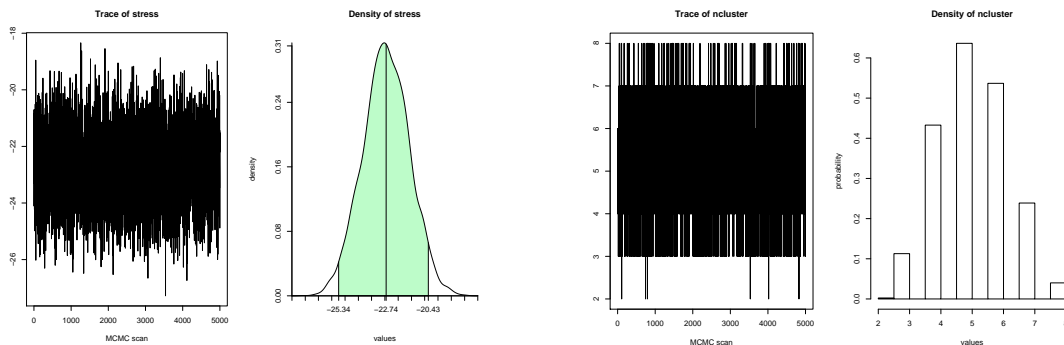
We run the algorithm for 57,000 iterations, where the first `nburn = 7,000` iterations are discarded and we use a thinning of `nskip = 10` to reduce the autocorrelation of the Markov chain. In short, we have a final sample of size `nsave = 5,000`.

We can obtain statistics of the posterior estimates and diagnostic tests on the model with the command

```
### SUMMARY
summary(fit)
```

While we plot the trace and the posterior kernel density estimation of a given parameter of the model (see Figure 4.3.3) with the following command-line:

```
### PLOT
plot(fit, ask=FALSE, nfigr=1, nfigc=2, param="stress")
plot(fit, ask=FALSE, nfigr=1, nfigc=2, param="ncluster")
```



(a) Traces and posterior kernel density estimation of  $\beta_1$  (b) Traces and posterior number of clusters  $K_J$

Figure 4.17: Markov chain sample of  $\beta_1$  and of the number of clusters  $K_J$

## 4.4 Conclusions

We briefly summarize and compare the results obtained for the different models.

The first model with nonparametric error does not improve remarkably the estimates with respect to the parametric model of Leon et al. [24]. We recall that Leon et al. [24]’s estimates are affected by a large Monte Carlo error and so they are not reliable. The interval estimates of the median failure time at stress level 22.5MPa are robust and comparable to those of the parametric model; while the interval estimates of the 1st percentile of the failure time distribution at stress level 23.4MPa are less robust.

The interval estimates of the quantiles of interest under the second model (with nonparametric random-effect) are narrower with respect to those of the first model. The nonparametric prior distribution of the random-effects is more flexible with respect to the parametric one of the first model, and so we better estimate the spool-effects. Nonetheless, both the credibility intervals are quite sensitive to the choice of the hyperparameters of the NGG process prior.

The results for the model with Gamma-distributed error, that we fitted with *DPpackage*, support those of the second model. Also in this case, the estimates of the quantiles of interest are sensitive to the choice of the hyperparameters of the prior.

Summing up, the model with nonparametric random-effect improves the predictive estimates with respect to the benchmark model of Leon et al. [24], but the interval estimates are still too wide to make statements about the reliability of the pressure vessels of the Space Shuttle. We believe that this uncertainty is due to the remarkable variability among the samples spools and to the small sample size (only 8 spools) from the potentially infinite population of spools.

## Appendix A

# Full-conditionals of the nonparametric error model

Here we compute the full-conditionals of the Gibb-sampler of the model with nonparametric error. First, we compute the full-conditionals of the parametric part of the model. In particular, we show that the full-conditionals of  $\beta_0$ ,  $\beta_1$  and  $\alpha_j$ , for  $j = 1, \dots, J$ , have log-concave density distribution functions; while the full-conditionals of  $\lambda$  and  $\alpha_{\text{new}}$  are conjugate. Then, we provide the full-conditionals of the nonparametric part of the model.

### A.1 Parametric full-conditionals

The full-conditionals of  $\beta_0$ ,  $\beta_1$  and  $\alpha_j$ , for  $j = 1, \dots, J$  are log-concave and hence we can sample from them using an acceptance rejection sampling method. The full-conditional of  $\beta_0$  has the following expression:

$$\begin{aligned}
 & \mathcal{L}(d\beta_0 | \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\
 &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(d\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(d\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)} \\
 &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)} \\
 &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)} \\
 &\propto \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_0) \\
 &\propto \prod_{i=1}^n \{f(t_i | \beta_0, \beta_1, \alpha_{k[i]}, \theta_i)\} \mathcal{L}(d\beta_0) \\
 &\propto \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\beta_0^2}{2\sigma_0^2}\right) \prod_{i=1}^n \exp\left\{-\left(\frac{t_i}{\eta_i}\right)^{\theta_i}\right\} \frac{\theta_i}{\eta_i} \left(\frac{t_i}{\eta_i}\right)^{\theta_i-1} d\beta_0,
 \end{aligned}$$

where  $\eta_i = \exp\{\beta_0 + \beta_1 \log(x_i) + \alpha_{k[i]}\}$  is the linear predictor,  $k[i] = 1, \dots, J$  identifies the spool of observation  $i$ , and  $i = 1, \dots, n$ .

After some algebra we obtain

$$\begin{aligned} & \mathcal{L}(d\beta_0 | \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\ & \propto \exp\left\{ -\frac{\beta_0^2}{2\sigma_0^2} - \sum_{i=1}^n [t_i \exp\{-\beta_0 - \beta_1 \log(x_{s,i}) - \alpha_{k[i]}\}]^{\theta_i} - \beta_0 \sum_{i=1}^n \theta_i \right\} d\beta_0. \end{aligned}$$

Similarly, we find the expression of the full-conditional of  $\beta_1$

$$\begin{aligned} & \mathcal{L}(d\beta_1 | \beta_0, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\ & = \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, d\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, d\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)} \\ & = \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0, d\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0, d\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)} \\ & = \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(d\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(d\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)} \\ & \propto \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\beta_1) \\ & \propto \prod_{i=1}^n \{f(t_i | \beta_0, \beta_1, \alpha_{k[i]}, \theta_i)\} \mathcal{L}(d\beta_1) \\ & \propto \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{\beta_1^2}{2\sigma_1^2}\right) \prod_{i=1}^n \exp\left\{-\left(\frac{t_i}{\eta_i}\right)^{\theta_i}\right\} \frac{\theta_i}{\eta_i} \left(\frac{t_i}{\eta_i}\right)^{\theta_i-1} d\beta_1 \\ & \propto \exp\left\{-\frac{\beta_1^2}{2\sigma_1^2} - \sum_{i=1}^n [t_i \exp\{-\beta_0 - \beta_1 \log(x_{s,i}) - \alpha_{k[i]}\}]^{\theta_i} - \beta_1 \sum_{i=1}^n \log(x_{s,i}) \theta_i\right\} d\beta_1. \end{aligned}$$

For  $j = 1, \dots, J$ , the full-conditional of  $\alpha_j$  has the following density distribution func-

tion

$$\begin{aligned}
& \mathcal{L}(d\alpha_j | \beta_0, \beta_1, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}}, \lambda)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}} | \lambda) \mathcal{L}(\lambda)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \alpha_{\text{new}} | \lambda) \mathcal{L}(\lambda)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j | \lambda) \mathcal{L}(\boldsymbol{\alpha}_{-j}, \alpha_{\text{new}} | \lambda) \mathcal{L}(\lambda)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j | \lambda) \mathcal{L}(\boldsymbol{\alpha}_{-j}, \alpha_{\text{new}} | \lambda) \mathcal{L}(\lambda)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j | \lambda)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j | \lambda)} \\
&\propto \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\alpha_j | \lambda) \\
&\propto \prod_{i=1}^n \{f(t_i | \beta_0, \beta_1, \alpha_{k[i]}, \theta_i)\} \mathcal{L}(d\alpha_j | \lambda) \\
&\propto \prod_{i:k[i]=j} \{f(t_i | \beta_0, \beta_1, \alpha_j, \theta_i)\} \mathcal{L}(d\alpha_j | \lambda) \\
&\propto \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{\alpha_j^2}{2\lambda}\right) \prod_{i:k[i]=j} \exp\left\{-\left(\frac{t_i}{\eta_i}\right)^{\theta_i}\right\} \frac{\theta_i}{\eta_i} \left(\frac{t_i}{\eta_i}\right)^{\theta_i-1} d\alpha_j, \\
&\propto \exp\left\{-\frac{\alpha_j^2}{2\lambda} - \sum_{i:k[i]=j} [t_i \exp\{-\beta_0 - \beta_1 \log(x_{s,i}) - \alpha_j\}]^{\theta_i} - \alpha_j \sum_{i:k[i]=j} \theta_i\right\} d\alpha_j.
\end{aligned}$$

Now we show that the full-conditional of  $\lambda$  has Inverse Gamma distribution:

$$\begin{aligned}
& \mathcal{L}(d\lambda | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda, P, \boldsymbol{\theta}, U)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda) \mathcal{L}(P, \boldsymbol{\theta}, U)} \\
&= \frac{\mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, d\lambda)} \\
&= \frac{\mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}} | \lambda) \mathcal{L}(d\lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}} | \lambda) \mathcal{L}(d\lambda)} \\
&\propto \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}} | \lambda) \mathcal{L}(d\lambda)
\end{aligned}$$

We recall that  $\mathcal{L}(\alpha_j | \lambda) \stackrel{\text{iid}}{\sim} N(0, \lambda)$ , for  $j = 1, \dots, J, (J + 1 = \text{new})$ , and  $\mathcal{L}(\lambda) \sim$



$InvGamma(a_0, b_0)$ . And so,

$$\begin{aligned}
& \mathcal{L}(d\lambda|\beta_0, \beta_1, \boldsymbol{\alpha}, P, U, \mathbf{T} = \mathbf{t}) \\
& \propto \prod_{j=1}^{J+1} \mathcal{L}(\alpha_j|\lambda)\mathcal{L}(d\lambda). \\
& \propto \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{-a_0-1} \exp\left\{-\frac{b_0}{\lambda}\right\} \prod_{j=1}^{J+1} \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{\alpha_j^2}{2\lambda}\right\} d\lambda \\
& \propto \lambda^{-a_0-J/2-1} \exp\left\{-\frac{1}{\lambda}\left(\frac{1}{2}\sum_{j=1}^{J+1} \alpha_j^2 + b_0\right)\right\} d\lambda,
\end{aligned}$$

if  $\lambda > 0$ , 0 otherwise. It follows that

$$\mathcal{L}(d\lambda|\beta_0, \beta_1, \boldsymbol{\alpha}, P, U, \mathbf{T} = \mathbf{t}) \sim InvGamma\left(a_0 + \frac{J+1}{2}, b_0 + \frac{\sum_{j=1}^J \alpha_j^2 + \alpha_{new}^2}{2}\right)$$

Finally, we show that also the full-conditional of  $\alpha_{new}$  is conjugate:

$$\begin{aligned}
& \mathcal{L}(d\alpha_{new}|\beta_0, \beta_1, \boldsymbol{\alpha}, \lambda, P, \boldsymbol{\theta}, U, \mathbf{T} = \mathbf{t}) \\
& = \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{new}, \lambda, P, \boldsymbol{\theta}, U)\mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, d\alpha_{new}, \lambda, P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{new}, \lambda, P, \boldsymbol{\theta}, U)\mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, d\alpha_{new}, \lambda, P, \boldsymbol{\theta}, U)} \\
& = \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta})\mathcal{L}(\beta_0)\mathcal{L}(\beta_1)\mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}, \lambda)\mathcal{L}(P, \boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta})\mathcal{L}(\beta_0)\mathcal{L}(\beta_1)\mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}, \lambda)\mathcal{L}(P, \boldsymbol{\theta}, U)} \\
& = \frac{\mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}, \lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}, \lambda)} \\
& = \frac{\mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}|\lambda)\mathcal{L}(\lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}, d\alpha_{new}|\lambda)\mathcal{L}(\lambda)} \\
& = \frac{\mathcal{L}(\boldsymbol{\alpha}|\lambda)\mathcal{L}(d\alpha_{new}|\lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}|\lambda)\mathcal{L}(d\alpha_{new}|\lambda)} \\
& = \mathcal{L}(d\alpha_{new}|\lambda) \sim N(0, \lambda).
\end{aligned}$$

## A.2 Full-conditionals of the error

Here we provide the full-conditionals of  $P$ ,  $\boldsymbol{\theta}$  and  $U$ . As we mentioned in (3.13), the

full-conditional of  $P$  depends only by  $\boldsymbol{\theta}$  and  $U$ . By (1.13), we know that

$$\mathcal{L}(P|\boldsymbol{\theta}, U = u) = \frac{1}{\sum_{l=1}^{\infty} J_{l,u} + \sum_{l=1}^k L_l} \sum_{l=1}^{\infty} J_{l,u} \delta_{\tau_l} + \frac{1}{\sum_{l=1}^{\infty} J_{l,u} + \sum_{l=1}^k L_l} \sum_{l=1}^k L_l \delta_{\psi_l},$$

where  $(J_{l,u})_{l \geq 1}$  are the ranked points of a Poisson process on  $\mathbb{R}^+$  with mean intensity  $\rho(ds)$  (see (1.5) for the definition of the mean intensity measure);  $\tau_l \stackrel{iid}{\sim} P_0$ ;  $L_l \sim Gamma(e_l -$

$\sigma, \omega + u$ );  $\psi_1, \psi_2, \dots, \psi_k$  are the observed distinct values among  $\boldsymbol{\theta}$ ;  $e_l = \#\{X_i : X_i = \psi_l, 1 \leq i \leq n\}$ ; and  $l = 1, \dots, k$ .

As we mentioned in Section 3.3.2, we will sample only a finite number of jumps  $J_{l,u}$  and locations  $\tau_l$ . So, we compute the integer  $L$  such that the finite series stopped at  $L$  guarantees a good approximation of the infinite series through the criterion proposed by Argiento et al. [1].

It is trivial to sample the jumps  $\tau_l$ , while to sample the jumps  $J_{l,u}$  we use the method proposed by Ferguson and Klass [12]. Let  $(\varphi_l)_{l \geq 1}$  be a sequence of points from a homogeneous Poisson process with intensity equal to one, and  $R(x) = \int_x^{+\infty} \rho(ds) = (\kappa \omega^\sigma / \Gamma(1 - \sigma)) \Gamma(-\sigma; \omega x)$ . It can be proved that

$$J_{l,u} = R^{-1}(\varphi_l) = \frac{1}{\omega} \Gamma^{-1}\left(-\sigma; \frac{\varphi_l \Gamma(1 - \sigma)}{\omega^\sigma \kappa}\right), \quad l = 1, 2, \dots,$$

where  $\Gamma^{-1}$  is the inverse of the function  $x \mapsto \Gamma(-\sigma, x)$ . Notice that  $(\varphi_l)_{l \geq 1}$  can be obtained by the cumulative sum of i.i.d. exponential random variables with intensity equal to one.

We compute the full-conditional of  $\boldsymbol{\theta}$  in two successive steps applying the idea of Algorithm 2 of Neal [30]. First, we update the random partition  $\pi_n$  of  $\boldsymbol{\theta}$ , then, given the new partition, we update the distinct values  $\psi_1, \psi_2, \dots, \psi_k$  with the so-called *acceleration step*. This second step speeds up the mixing of the Markov chain.

$$\begin{aligned} & \mathcal{L}(d\boldsymbol{\theta} | \beta_0, \beta_1, \boldsymbol{\alpha}, P, U, \mathbf{T} = \mathbf{t}) \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, \lambda, P, d\boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \lambda, P, \boldsymbol{\theta}, U) \mathcal{L}(\beta_0, \beta_1, \boldsymbol{\alpha}, \lambda, P, d\boldsymbol{\theta}, U)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \lambda) \mathcal{L}(P, d\boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(\beta_0) \mathcal{L}(\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \lambda) \mathcal{L}(P, d\boldsymbol{\theta}, U)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(P, d\boldsymbol{\theta}, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(P, d\boldsymbol{\theta}, U)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\boldsymbol{\theta} | P) \mathcal{L}(P, U)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\boldsymbol{\theta} | P) \mathcal{L}(P, U)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\boldsymbol{\theta} | P)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\boldsymbol{\theta} | P)} \\ &\propto \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\theta}) \mathcal{L}(d\boldsymbol{\theta} | P) \\ &\propto \prod_{i=1}^n \mathcal{L}(T_i = t_i | \beta_0, \beta_1, \boldsymbol{\alpha}_{k[i]}, \theta_i) \mathcal{L}(d\theta_i | P). \end{aligned} \tag{A.1}$$

By (A.1),  $\theta_1, \dots, \theta_n$  are independent, given  $\beta_0, \beta_1, \boldsymbol{\alpha}$  and  $P$ .

$$\begin{aligned} & \mathcal{L}(d\boldsymbol{\theta}|\beta_0, \beta_1, \boldsymbol{\alpha}, P, U, \mathbf{T} = \mathbf{t}) \\ & \propto \prod_{i=1}^n \mathcal{L}(d\theta_i|\beta_0, \beta_1, \boldsymbol{\alpha}, P). \end{aligned}$$

Hence, we can update the  $n$  elements of the partition  $\pi_n$  one by one. Notice that

$$\begin{aligned} & \mathcal{L}(d\theta_i|P) \sim P \\ & P \simeq \sum_{l=1}^L p_{l,u} \delta_{\tau_l} + \sum_{l=1}^k p_l \delta_{\psi_l}, \end{aligned}$$

for  $i = 1, \dots, n$ , where  $p_{l,u} := J_{l,u}/(\sum_{l=1}^L J_{l,u} + \sum_{l=1}^k L_l)$  and  $p_l := L_l/(\sum_{l=1}^L J_{l,u} + \sum_{l=1}^k L_l)$ .

Hence the conditional law of  $\theta_i$  has the following discrete distribution

$$\mathcal{L}(d\theta_i|\beta_0, \beta_1, \boldsymbol{\alpha}, P) \propto \left\{ \sum_{l=1}^L p_{l,u} \delta_{\tau_l} + \sum_{l=1}^k p_l \delta_{\psi_l} \right\} f(t_i|\beta_0, \beta_1, \alpha_{k[i]}, \theta_i) \mathcal{L}(d\theta_i).$$

Observe that we are only interested in the new configuration  $\pi_n$ , i.e. the new grouping among  $\theta_i$ 's, and not in the  $\theta_i$ 's values.

Now, given the new configuration  $\pi_n$  and the other parameters, we update the distinct values  $\psi_1, \psi_2, \dots, \psi_k$ . Using (1.8), it can be proved that the full-conditionals of  $\boldsymbol{\psi}$  has the following expression:

$$\mathcal{L}(\boldsymbol{\psi}|\pi_n, \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T} = \mathbf{t}) = \prod_{l=1}^k \prod_{i:\pi_n(i)=l} f(t_i|\beta_0, \beta_1, \alpha_{k[i]}, \psi_l) P_0(d\psi_l),$$

where  $P_0 \sim \text{Gamma}(a_0, b_0)$  and the notation  $i : \pi_n(i) = l$  represents the set of the observations which have  $\theta_i = \psi_l$ . Therefore the  $\psi_l$ 's are reciprocally independent and each full-conditional of  $\psi_l$  has the following density distribution function

$$\mathcal{L}(\psi_l|\pi_n, \beta_0, \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \lambda, P, U, \mathbf{T} = \mathbf{t}) = \prod_{i:\pi_n(i)=l} f(t_i|\beta_0, \beta_1, \alpha_{k[i]}, \psi_l) P_0(d\psi_l), \quad (\text{A.2})$$

To sample from (A.2), we use a step of Metropolis-Hastings.

Finally, James et al. [18] proved that the full-conditional of the nuisance parameter  $U$  depends only by  $\boldsymbol{\theta}$  and has the following density distribution function

$$f(u|\boldsymbol{\theta}) = (u + w)^{k\sigma - n} u^{n-1} \exp\left[-\frac{\kappa}{\sigma}(u + w)^\sigma\right], \quad u > 0.$$

## Appendix B

# Full-conditionals of the nonparametric random effects model

Here we provide a factorization of the posterior distribution of the nonparametric random effects model, and then we derive the full-conditionals of the Gibb-sampler.

$$\begin{aligned}
& \mathcal{L}(\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta | \mathbf{T} = \mathbf{t}) \\
&= \frac{\mathcal{L}(\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta, \mathbf{T} = \mathbf{t})}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta) \mathcal{L}(\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta)}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(\beta_1, \boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda, \theta)}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(\beta_1) \mathcal{L}(\theta) \mathcal{L}(\boldsymbol{\alpha}, \alpha_{\text{new}}, \mu, \lambda) \mathcal{L}(\theta)}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(\beta_1) \mathcal{L}(\theta) \mathcal{L}(\alpha_{\text{new}} | \boldsymbol{\alpha}, \mu, \lambda) \mathcal{L}(\boldsymbol{\alpha}, \mu, \lambda) \mathcal{L}(\theta)}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \mathcal{L}(\alpha_{\text{new}} | \boldsymbol{\alpha}, \mu, \lambda) \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(\beta_1) \mathcal{L}(\theta) \mathcal{L}(\boldsymbol{\alpha}, \mu, \lambda) \mathcal{L}(\theta)}{\mathcal{L}(\mathbf{T} = \mathbf{t})} \\
&= \mathcal{L}(\alpha_{\text{new}} | \boldsymbol{\alpha}, \mu, \lambda) \mathcal{L}(\beta_1, \boldsymbol{\alpha}, \mu, \lambda, \theta | \mathbf{T} = \mathbf{t})
\end{aligned}$$

First, we compute the full-conditionals of the parametric part of the model. In particular, we show that the full-conditionals of  $\beta_1$  and  $\theta$  have log-concave density distribution functions. Then, we provide the full-conditionals of the nonparametric part of the model: the full-conditionals of  $\mu$  and  $\lambda$  are conjugate; while we use the Polya-urn scheme to sample from the full-conditionals of  $\alpha_1, \dots, \alpha_J$ . Finally we provide the expression of  $\alpha_{\text{new}} | \boldsymbol{\alpha}, \mu, \lambda$ .

## B.1 Parametric full-conditionals

The full-conditionals of  $\beta_1$  and  $\theta$  are log-concave and hence we can sample from them using an acceptance rejection sampling method. The full-conditional of  $\beta_1$  has the following expression:

$$\begin{aligned}
& \mathcal{L}(d\beta_1 | \boldsymbol{\alpha}, \lambda, \mu, \theta, \mathbf{T} = \mathbf{t}) \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta) \mathcal{L}(d\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta) \mathcal{L}(d\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(d\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(d\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(d\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \lambda, \mu) \mathcal{L}(\theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(d\beta_1) \mathcal{L}(\boldsymbol{\alpha}, \lambda, \mu) \mathcal{L}(\theta)} \\
&\propto \mathcal{L}(\mathbf{T} = \mathbf{t} | \beta_1, \boldsymbol{\alpha}, \theta) \mathcal{L}(d\beta_1) \\
&\propto \prod_{i=1}^n \{f(t_i | \beta_1, \alpha_{k[i]}, \theta)\} \mathcal{L}(d\beta_1) \\
&\propto \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{\beta_1^2}{2\sigma_1^2}\right) \prod_{i=1}^n \exp\left\{-\left(\frac{t_i}{\eta_i}\right)^\theta\right\} \frac{\theta}{\eta_i} \left(\frac{t_i}{\eta_i}\right)^{\theta-1} d\beta_1 \\
&\propto \exp\left\{-\frac{\beta_1^2}{2\sigma_1^2} - \sum_{i=1}^n [t_i \exp\{-\beta_1 \log(x_{s,i}) - \alpha_{k[i]}\}]^\theta - \beta_1 \sum_{i=1}^n \log(x_{s,i}) \theta\right\} d\beta_1,
\end{aligned}$$

where  $\eta_i = \exp\{\beta_0 + \beta_1 \log(x_i) + \alpha_{k[i]}\}$  is the linear predictor,  $k[i] = 1, \dots, J$  identifies the spool of observation  $i$ , and  $i = 1, \dots, n$ .

The full-conditional of  $\theta$  has the following density distribution function:

$$\begin{aligned}
& \mathcal{L}(d\theta|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \mathbf{T} = \mathbf{t}) \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)\mathcal{L}(\beta_1, \boldsymbol{\alpha}, \lambda, \mu, d\theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)\mathcal{L}(\beta_1, \boldsymbol{\alpha}, \lambda, \mu, d\theta)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(\beta_1, \boldsymbol{\alpha}, \lambda, \mu, d\theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(\beta_1, \boldsymbol{\alpha}, \lambda, \mu, d\theta)} \\
&= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(\beta_1)\mathcal{L}(\boldsymbol{\alpha}, \lambda, \mu)\mathcal{L}(\theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\beta_1)\mathcal{L}(\boldsymbol{\alpha}, \lambda, \mu)\mathcal{L}(\theta)} \\
&\propto \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\theta) \\
&\propto \prod_{i=1}^n \{f(t_i|\beta_1, \alpha_{k[i]}, \theta)\} \mathcal{L}(d\theta) \\
&\propto \theta^{a_0-1} \exp(-b_0\theta) \prod_{i=1}^n \exp\left\{-\left(\frac{t_i}{\eta_i}\right)^\theta\right\} \frac{\theta}{\eta_i} \left(\frac{t_i}{\eta_i}\right)^{\theta-1} d\theta \\
&\propto \exp\left\{(a_0 + n - 1) \log(\theta) + \theta \left[ \sum_{i=1}^n (\log(t_i) - \beta_1 \log(x_{s,i}) - \alpha_{k[i]}) - b_0 \right] - \right. \\
&\quad \left. - \sum_{i=1}^n [t_i \exp\{-\beta_1 \log(x_{s,i}) - \alpha_{k[i]}\}]^\theta \right\} d\theta,
\end{aligned}$$

if  $\theta > 0$ , 0 otherwise.

## B.2 Full-conditionals of the random effects

First we show that the full-conditionals of  $\mu$  and  $\lambda$  are conjugate; then we provide the Polya-urn scheme of the full-conditionals of  $\alpha_j$ 's, where  $j = 1, \dots, J$ . Finally, we derive the expression of  $\alpha_{\text{new}}|\boldsymbol{\alpha}, \mu, \lambda$ .

Notice that both the full-conditionals do not depend directly on the failure times  $\mathbf{T}$ , since

$$\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta) = \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta).$$

Let start with the full-conditional of  $\mu$ :

$$\begin{aligned}
& \mathcal{L}(d\mu|\beta_1, \boldsymbol{\alpha}, \theta, \lambda, \mathbf{T} = \mathbf{t}) = \mathcal{L}(d\mu|\beta_1, \boldsymbol{\alpha}, \theta, \lambda) \\
&= \frac{\mathcal{L}(d\mu, \beta_1, \boldsymbol{\alpha}, \theta, \lambda)}{\int \mathcal{L}(d\mu, \beta_1, \boldsymbol{\alpha}, \theta, \lambda)} = \frac{\mathcal{L}(d\mu, \boldsymbol{\alpha}, \lambda)\mathcal{L}(\beta_1, \theta)}{\int \mathcal{L}(d\mu, \boldsymbol{\alpha}, \lambda)\mathcal{L}(\beta_1, \theta)} \\
&= \frac{\mathcal{L}(d\mu, \boldsymbol{\alpha}, \lambda)}{\int \mathcal{L}(d\mu, \boldsymbol{\alpha}, \lambda)} = \frac{\mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda)\mathcal{L}(d\mu)\mathcal{L}(\lambda)}{\int \mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda)\mathcal{L}(d\mu)\mathcal{L}(\lambda)} \\
&\propto \mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda)\mathcal{L}(d\mu).
\end{aligned}$$

By (1.8), we know that

$$\mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda) = p(e_1, \dots, e_k) \prod_{l=1}^k P_0(\psi_l|\mu, \lambda),$$

where  $P_0(\cdot|\mu, \lambda) \sim N(\mu, \lambda)$  and  $\psi_1, \dots, \psi_k$  are the distinct values among  $\alpha_1, \dots, \alpha_J$ . Hence

$$\begin{aligned} & \mathcal{L}(d\mu|\beta_1, \boldsymbol{\alpha}, \theta, \lambda, \mathbf{T} = \mathbf{t}) \\ & \propto p(e_1, \dots, e_k) \prod_{l=1}^k P_0(\psi_l|\mu, \lambda) \mathcal{L}(d\mu) \\ & \propto \prod_{l=1}^k P_0(\psi_l|\mu, \lambda) \mathcal{L}(d\mu) \\ & \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{\mu^2}{2\sigma_0^2}\right] \prod_{l=1}^k \left\{ \frac{1}{\sqrt{2\lambda}} \exp\left[-\frac{(\psi_l - \mu)^2}{2\lambda}\right] \right\} d\mu \\ & \propto \exp\left\{ -\mu^2 \left(\frac{k}{2\lambda} + \frac{1}{2\sigma_0^2}\right) + \mu \frac{\sum_{l=1}^k \psi_l}{\lambda} - \frac{\sum_{l=1}^k \psi_l^2}{2\lambda} \right\} d\mu. \end{aligned}$$

After some algebra, we obtain that

$$\mathcal{L}(d\mu|\beta_1, \boldsymbol{\alpha}, \theta, \lambda, \mathbf{T} = \mathbf{t}) \sim N\left(\frac{\sum_{l=1}^k \psi_l}{\lambda} \left(\frac{k}{\lambda} + \frac{1}{\sigma_0^2}\right)^{-1}, \left(\frac{k}{\lambda} + \frac{1}{\sigma_0^2}\right)^{-1}\right).$$

Similarly, we can compute the full-conditional of  $\lambda$ :

$$\begin{aligned} & \mathcal{L}(d\lambda|\beta_1, \boldsymbol{\alpha}, \theta, \mu, \mathbf{T} = \mathbf{t}) = \mathcal{L}(d\lambda|\beta_1, \boldsymbol{\alpha}, \theta, \mu) \\ & = \frac{\mathcal{L}(d\lambda, \beta_1, \boldsymbol{\alpha}, \theta, \mu)}{\int \mathcal{L}(d\lambda, \beta_1, \boldsymbol{\alpha}, \theta, \mu)} = \frac{\mathcal{L}(d\lambda, \boldsymbol{\alpha}, \mu) \mathcal{L}(\beta_1, \theta)}{\int \mathcal{L}(d\lambda, \boldsymbol{\alpha}, \mu) \mathcal{L}(\beta_1, \theta)} \\ & = \frac{\mathcal{L}(d\lambda, \boldsymbol{\alpha}, \mu)}{\int \mathcal{L}(d\lambda, \boldsymbol{\alpha}, \mu)} = \frac{\mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda) \mathcal{L}(d\lambda) \mathcal{L}(\mu)}{\int \mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda) \mathcal{L}(d\lambda) \mathcal{L}(\mu)} \\ & \propto \mathcal{L}(\boldsymbol{\alpha}|\mu, \lambda) \mathcal{L}(d\lambda) \\ & \propto p(e_1, \dots, e_k) \prod_{l=1}^k P_0(\psi_l|\mu, \lambda) \mathcal{L}(d\lambda) \tag{B.1} \\ & \propto \prod_{l=1}^k P_0(\psi_l|\mu, \lambda) \mathcal{L}(d\lambda) \\ & \propto \frac{(\tau_2/2)^{(\tau_1/2)}}{\Gamma(\tau_1/2)} \lambda^{-\tau_1/2-1} \exp\left[-\frac{\tau_2}{2\lambda}\right] \prod_{l=1}^k \left\{ \frac{1}{\sqrt{2\lambda}} \exp\left[-\frac{(\psi_l - \mu)^2}{2\lambda}\right] \right\} d\lambda \\ & \propto \lambda^{-\tau_1/2-k/2-1} \exp\left[-\frac{\sum_{l=1}^k (\psi_l - \mu)^2 - \tau_2}{2\lambda}\right] d\lambda, \end{aligned}$$

if  $\lambda > 0$ , 0 otherwise. Notice the kernel of the Inverse Gamma in the last line of (B.1), so

$$\mathcal{L}(d\lambda|\beta_1, \boldsymbol{\alpha}, \theta, \mu, \mathbf{T} = \mathbf{t}) \sim \text{InvGamma}\left(\frac{\tau_1 + k}{2}, \frac{\sum_{l=1}^k (\psi_l - \mu)^2 + \tau_2}{2}\right).$$

For  $j = 1, \dots, J$ , the full-conditional of  $\alpha_j$  has the following density distribution function

$$\begin{aligned} & \mathcal{L}(d\alpha_j|\beta_1, \boldsymbol{\alpha}_{-j}, \lambda, \mu, \theta, \mathbf{T} = \mathbf{t}) \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)\mathcal{L}(\beta_1, d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu, \theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \lambda, \mu, \theta)\mathcal{L}(\beta_1, d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu, \theta)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(\beta_1)\mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu)\mathcal{L}(\theta)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(\beta_1)\mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu)\mathcal{L}(\theta)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j, \boldsymbol{\alpha}_{-j}, \lambda, \mu)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu)\mathcal{L}(\boldsymbol{\alpha}_{-j}, \lambda, \mu)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu)\mathcal{L}(\boldsymbol{\alpha}_{-j}, \lambda, \mu)} \\ &= \frac{\mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu)}{\int \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu)} \\ &\propto \mathcal{L}(\mathbf{T} = \mathbf{t}|\beta_1, \boldsymbol{\alpha}, \theta)\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu) \\ &\propto \prod_{i=1}^n \{f(t_i|\beta_1, \alpha_{k[i]}, \theta)\} \mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu) \\ &\propto \prod_{i:k[i]=j} \{f(t_i|\beta_1, \alpha_j, \theta)\} \mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu). \end{aligned}$$

By (1.10), we know that

$$\mathcal{L}(d\alpha_j|\boldsymbol{\alpha}_{-j}, \lambda, \mu) = w_0(J-1, k^{-j}; \sigma, \kappa)P_0(d\alpha_j) + w_1(J-1, k^{-j}; \sigma, \kappa) \sum_{l=1}^{k^{-j}} (e_l^{-j} - \sigma)\delta_{\psi_l}(d\alpha_j),$$

where  $w_0(J-1, k^{-j}; \sigma, \kappa)$  and  $w_1(J-1, k^{-j}; \sigma, \kappa)$  are defined in (1.11), while  $k^{-j}$  represents the number of distinct values among  $\boldsymbol{\alpha}_{-j}$ . Hence

$$\mathcal{L}(d\alpha_j|\beta_1, \boldsymbol{\alpha}_{-j}, \lambda, \mu, \theta, \mathbf{T} = \mathbf{t}) = \frac{\tilde{w}_0(\alpha_j)P_0(d\alpha_j) + \sum_{l=1}^{k^{-j}} \tilde{w}_l(\psi_l)\delta_{\psi_l}(d\alpha_j)}{\int \tilde{w}_0(\alpha_j)P_0(d\alpha_j) + \sum_{l=1}^{k^{-j}} \tilde{w}_l(\psi_l)} \quad (\text{B.2})$$

where  $\tilde{w}_0(\alpha_j) := w_0(J-1, k^{-j}; \sigma, \kappa) \prod_{i:k[i]=j} f(t_i|\beta_1, \alpha_j, \theta)$  and  $\tilde{w}_l(\psi_l) := w_1(J-1, k^{-j}; \sigma, \kappa)(e_l^{-j} - \sigma) \prod_{i:k[i]=j} f(t_i|\beta_1, \psi_l, \theta)$ , for  $l = 1, \dots, k^{-j}$ .

We do not have any analytic expression of the integral  $\int \tilde{w}_0(\alpha_j)P_0(d\alpha_j)$  in (B.2). To get over the evaluation of this integral, we compute the full-conditionals of  $\alpha_j$ 's in two successive steps using Algorithm 8 of Neal [30] for non-conjugated priors. First, we update the random partition  $\pi_J$  of  $\boldsymbol{\alpha}$  with the support of a nuisance variable, then, given the new random partition, we update the distinct values  $\psi_1, \psi_2, \dots, \psi_k$  with the so-called



*acceleration step.* This second step of the algorithm speeds up the mixing of the Markov chain.

We introduce some notation before providing the steps of the algorithm. To a given random partition  $\pi_J$  there exists one and only one configuration vector  $\mathbf{c} = (c_1, \dots, c_J)$ , where  $c_j \in \{1, \dots, k\}$ ,  $k$  is the number of distinct values  $\psi_1, \psi_2, \dots, \psi_k$  and  $c_j$  identifies the value of  $\alpha_j$ , for  $j = 1, \dots, J$ . For instance,  $c_2 = 3$  iff  $\alpha_2 = \psi_3$ . While  $\mathbf{c}_{-j}$  represents the vector  $\mathbf{c}$  not including the  $j$ -th component, and  $C_{-j} := \{c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_J\}$ .

Here we update the configuration vector  $\mathbf{c}$ .

For  $j = 1, \dots, J$

- if  $c_j = c$ , for some  $c \in C_{-j}$ 
  - draw  $\psi_{k+1} \sim P_0$
  - draw a new value for  $c_j$  from  $\{1, \dots, k+1\}$  using the following probability distribution

$$\mathbb{P}(c_j = c | \mathbf{c}_{-j}, \boldsymbol{\psi}, \psi_{k+1}, \mu, \lambda, \beta_1, \theta, \mathbf{T} = \mathbf{t})$$

$$= \begin{cases} \omega_1(J, k+1)(e_c^{-j} - \sigma) \prod_{i:k[i]=j} f(t_i | \beta_1, \psi_c, \theta) & 1 \leq c \leq k \\ \omega_0(J, k+1) \prod_{i:k[i]=j} f(t_i | \beta_1, \psi_c, \theta) & c = k+1 \end{cases}$$

- if  $c_j \neq c$ , for any  $c \in C_{-j}$ 
  - draw a new value for  $c_j$  from  $\{1, \dots, k\}$  using the following probability distribution

$$\mathbb{P}(c_j = c | \mathbf{c}_{-j}, \boldsymbol{\psi}, \mu, \lambda, \beta_1, \theta, \mathbf{T} = \mathbf{t})$$

$$= \begin{cases} \omega_1(J, k)(e_c^{-j} - \sigma) \prod_{i:k[i]=j} f(t_i | \beta_1, \psi_c, \theta) & c \neq c_j \\ \omega_0(J, k) \prod_{i:k[i]=j} f(t_i | \beta_1, \psi_c, \theta) & c = c_j \end{cases}$$

Observe that we are only interested in the new configuration  $\mathbf{c}$ , i.e. the new grouping among  $\alpha_j$ 's, and not in the  $\alpha_j$ 's values.

Now we sample the new distinct values  $\psi_1, \dots, \psi_k$ , given the new configuration  $\mathbf{c}$  and the other parameters. Using (1.8), it can be proved that the full-conditionals of  $\boldsymbol{\psi}$  has the following expression:

$$\mathcal{L}(\boldsymbol{\psi} | \mathbf{c}, \beta_1, \boldsymbol{\alpha}, \mu, \lambda, \mathbf{T} = \mathbf{t}) = \prod_{l=1}^k \prod_{i:\mathbf{c}[i]=l} f(t_i | \beta_1, \alpha_{k[i]}, \psi_l) P_0(d\psi_l),$$

where  $P_0 \sim N(\mu, \lambda)$ . Therefore the  $\psi_l$ 's are reciprocally independent and each full-conditional of  $\psi_l$  has the following density distribution function

$$\begin{aligned}
& \mathcal{L}(\psi_l | \mathbf{c}, \beta_1, \boldsymbol{\alpha}, \mu, \lambda, \mathbf{T} = \mathbf{t}) \\
&= \prod_{i: \mathbf{c}(k[i])=l} f(t_i | \beta_1, \psi_l, \theta) P_0(d\psi_l) \\
&\propto \exp \left\{ -\frac{(\psi_l - \mu)^2}{2\lambda} - \sum_{i: \mathbf{c}(k[i])=l} \psi_l \theta - \sum_{i: \mathbf{c}(k[i])=l} [t_i \exp(-\beta_1 \log(x_{s,i}) + \psi_l)]^\theta \right\}
\end{aligned}$$

Notice that the full-conditional of  $\psi_l$  has log-concave density distribution function, and so we can sample from it by an acceptance rejection sampling method.

Finally, the expression of  $\alpha_{\text{new}} | \boldsymbol{\alpha}, \mu, \lambda$  is trivial. In fact, by (1.10), we know that

$$\mathcal{L}(d\alpha_{\text{new}} | \boldsymbol{\alpha}, \lambda, \mu) = w_0(J, k; \sigma, \kappa) P_0(d\alpha_j) + w_1(J, k; \sigma, \kappa) \sum_{l=1}^k (e_l - \sigma) \delta_{\psi_l}(d\alpha_{\text{new}}),$$

where  $w_0(J, k; \sigma, \kappa)$  and  $w_1(J, k; \sigma, \kappa)$  are defined in (1.11).

# Appendix C

## Data-set

Table C.1: Failure-times of NASA pressure vessels wrapped with Kevlar. Right-censored observations are indicated with an asterisk \*.

Stress (MPa)	Spool	F-time (hours)	Stress (MPa)	Spool	F-time (hours)
29.7	2	2.2	29.7	5	243.9
29.7	7	4.0	29.7	4	254.1
29.7	7	4.0	29.7	1	444.4
29.7	7	4.6	29.7	8	590.4
29.7	7	6.1	29.7	8	638.2
29.7	6	6.7	29.7	1	755.2
29.7	7	7.9	29.7	1	952.2
29.7	5	8.3	29.7	1	1108.2
29.7	2	8.5	29.7	4	1148.5
29.7	2	9.1	29.7	4	1569.3
29.7	2	10.2	29.7	4	1750.6
29.7	3	12.5	29.7	4	1802.1
29.7	5	13.3	27.6	3	19.1
29.7	7	14.0	27.6	3	24.3
29.7	3	14.6	27.6	3	69.8
29.7	6	15.0	27.6	2	71.2
29.7	3	18.7	27.6	3	136.0
29.7	2	22.1	27.6	2	199.1
29.7	7	45.9	27.6	2	403.7
29.7	2	55.4	27.6	2	432.2
29.7	7	61.2	27.6	1	453.4

Continued on next page

Table C.1 – continued from previous page

Stress (MPa)	Spool	F-time (hours)	Stress (MPa)	Spool	F-time (hours)
29.7	5	87.5	27.6	2	514.1
29.7	8	98.2	27.6	6	514.2
29.7	3	101.0	27.6	6	541.6
29.7	2	111.4	27.6	2	544.9
29.7	6	144.0	27.6	8	554.2
29.7	2	158.7	27.6	1	664.5
27.6	2	694.1	25.5	1	11487.3
27.6	4	876.7	25.5	5	11727.1
27.6	1	930.4	25.5	4	13501.3
27.6	6	1254.9	25.5	1	14032.0
27.6	4	1275.6	25.5	4	29808.0
27.6	4	1536.8	25.5	1	31008.0
27.6	1	1755.5	23.4	7	4000.0
27.6	8	2046.2	23.4	7	5376.0
27.6	4	6177.5	23.4	6	7320.0
25.5	6	225.2	23.4	3	86161.0
25.5	7	503.6	23.4	5	9120.0
25.5	3	1087.7	23.4	2	14440.0
25.5	2	1134.3	23.4	6	16104.0
25.5	2	1824.3	23.4	5	20231.0
25.5	2	1920.1	23.4	6	20233.0
25.5	2	2383.0	23.4	5	35880.0
25.5	3	2442.5	23.4	1	41000.0*
25.5	8	2974.6	23.4	1	41000.0*
25.5	2	3708.9	23.4	1	41000.0*
25.5	8	4908.9	23.4	1	41000.0*
25.5	2	5556.0	23.4	4	41000.0*
25.5	6	6271.1	23.4	4	41000.0*
25.5	8	7332.0	23.4	4	41000.0*
25.5	8	7918.7	23.4	4	41000.0*
25.5	6	7996.0	23.4	8	41000.0*
25.5	8	9240.3	23.4	8	41000.0*
25.5	8	9973.0	23.4	8	41000.0*

# Bibliography

- [1] Argiento, R., Guglielmi, A. and Pievatolo, A. (2008). Bayesian density estimation and model selection using nonparametric hierarchical mixtures. Technical report no. 08.12-MI, CNR IMATI, Milano.
- [2] Argiento, R., Guglielmi, A. and Pievatolo, A. (2010). Mixed-effects modelling of Kevlar fibre failure times through Bayesian nonparametrics. In *Complex data modeling and computationally intensive statistical methods*, eds. Mantovan, P. and Secchi, P., Springer, New York.
- [3] Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, **31**, 929-953.
- [4] Chaloner, K. (1991). Bayesian residuals analysis in presence of censoring. *Biometrika*, **31**, 651-659.
- [5] Composite Pressure Vessel and Structure Summit (2009). NASA White Sands Test Facility, Las Cruces, New Mexico.  
<http://www.nasa.gov/centers/wstf/news/safetysummit2009.html>
- [6] Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). *Statistical analysis of reliability data*, Chapman and Hall, London.
- [7] Dunson, D. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics*, ed. Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G., Cambridge University Press, Cambridge.
- [8] Escobar, M.D. (1994). Estimating Normal Means with a Dirichlet Process Prior. *Journal of the American Statistical Association*, **89**, 268-277.
- [9] Escobar, M.D. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, **90**, 577-588.
- [10] Feiveson, A.H. and Kulkarni, P.M. (2000). Reliability of Space-Shuttle pressure vessels with random batch effects. *Technometrics*, **42**, 332-344.

- [11] Ferguson, T.F. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209-230.
- [12] Ferguson, T.F. and Klass, M.J. (1972). A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics*, **43**, 1634-1643.
- [13] Gelman, A.(2005). Analysis of variance: why it is more important than ever. *Annals of Statistics*, **33**, 1-31.
- [14] Gerstle, F.P. and Kunz, S.C. (1983). Prediction of Long-Term Failure in Kevlar 49 Composites. In long behaviour composites. *American Society for Testing and Materials*, Philadelphia, 263-292.
- [15] Ghosh, J.K. and Ramamoorthi, R.V. (2003). *Bayesian Nonparametrics*, Springer, New York.
- [16] Glaser, R.E. (1983). *Statistical Analysis of Kevlar 49 Epoxy Composite Stress-Rupture Data*. Report UCID-19849, Lawrence Livermore National Laboratory.
- [17] Ishwaran, H. and James, L.F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statistica Sinica*, **13**, 1211-1236.
- [18] James, L.F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, **36**, 76-97.
- [19] Jara, A., Hanson, T.E., Quintana, F.A., Müller, P. and Rosner G.L. (2009). DP-package: Bayesian Non- and Semi-parametric Modelling in R. *Journal of Statistical Software*.
- [20] Kalbfleisch, J. and Prentice, R. (2001). *The statistical analysis of failure time data*, Wiley, New York.
- [21] Kalli, M., Griffin J.E. and Walker, S.G. (2009). Slice sampling mixture models *Statistics and computing*, DOI 10.1007/s11222-009-9150-y.
- [22] Kingman, J.F.C.(1993). *Poisson Processes*, Oxford Science Publications, Oxford.
- [23] Kleinman, K.P. and Ibrahim, J.G. (1998). A semi-parametric bayesian approach to generalized linear mixed models. *Statistics in Medicine*, **17**, 2579-2596.
- [24] Leon, R.V., Ashby, A.J. and Thyagarajan, J. (2006). Bayesian modeling of accelerated life tests with random effects. *Journal of quality technology*, **39**, 3-13.

- [25] Li, Y., Müller, P. and Lin, X. (2010). Center-adjusted inference for a nonparametric Bayesian random effect distribution. *Statistica Sinica*, to appear.
- [26] Lijoi, A., Mena, R. H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 715-740.
- [27] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*, ed. Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G., Cambridge University Press, Cambridge.
- [28] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, **12**, 351–357.
- [29] Müller, P. and Quintana, F.A. (2004). Nonparametric Bayesian Data Analysis. *Statistical Science*, **19**, 105-110.
- [30] Neal, R.M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, **9**, 249-265.
- [31] Pitman J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In: *Statistics, Probability and Game Theory (IMS Lecture Notes Monograph Series)*, **30**, 245-267.
- [32] Pitman J. (2003). Poisson-Kingman partitions. In D.R. Goldstein, editor, *Science and Statistics: a Festschrift for Terry Speed*, volume 40 of *Lectures Notes - Monograph Series*, 1-34. Institute of Mathematical Statistics, Hayward, California.
- [33] R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien.
- [34] Regazzini, E. (1996). Impostazione non parametrica di problemi di inferenza statistica bayesiana. Quaderni del Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma, La Sapienza.
- [35] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics*, **31**, 560–585.
- [36] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- [37] Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics-Simulation and Computation*, **36**, 4554.

- [38] Zeger, S.L. and Karim, M.R. (1991). Generalized Linear Models With Random Effects; A Gibbs Sampling Approach. *Journal of the American Statistical Association*, **86**, 79-86.