UNIVERSITAT DE VALÈNCIA

Programa de Doctorat en Estadística i Optimització

Sequential Monte Carlo methods in Bayesian joint models for longitudinal and time-to-event data

by

Danilo Alvares da Silva

Thesis submitted for the degree of Doctor of
Philosophy in Statistics and Optimisation

Supervised by
Carmen Armero i Cervera and
Anabel Forte Deltell

in the
Faculty of Mathematics
Department of Statistics and Operations Research

April 2017

Acknowledgements

Brasil

"Depois de 20 anos na escola" com certeza muita gente foi fundamental na minha formação acadêmica e pessoal, por isso serei eternamente grato a essas pessoas. Começando lá na creche da UNESP/Jaboticabal, passando pelas escolas Nossa Senhora Aparecida, SESI e ETE Dr. Adail Nunes da Silva, e Cursinho Ativo, minhas lembranças remetem sempre à bons professores e amigos. Externo ao contexto escolar, minha família tem sido o ponto chave em todas essas fases da minha vida.

Clarice Alvares, esse é o nome da guerreira que, junto com minha irmã Patrícia, me apoiou em tudo. Ainda na minha infância surgiu outro suporte de peso, Orlando, meu padrasto/pai/padrinho. Indiscutivelmente, madrinha/padrinho, irmãos, tia(o)s, prima(o)s, sobrinhas, avós e cunhado foram figuras fundamentais durante toda essa jornada. Meu pai e minha madrasta também foram pontos de equilíbrio nesse meu caminho.

Passado esse período de formação pessoal e escolar, meu primeiro grande desafio foi o curso de Ciência da Computação na Universidade Estadual de Londrina (UEL). Ali reencontrei alguns amigos, fiz novas amizades e tive novas perspectivas de vida. Na UEL despertei novos interesses acadêmicos e no final do meu primeiro ano universitário eu já estava decidido a prestar vestibular na Universidade de São Paulo (USP) para o curso de Matemática Aplicada e Computação Científica.

Meus anos de USP/São Carlos, ou simplesmente CAASO, foram fantásticos em todos os sentidos. Nesse período, além da graduação, também fiz intercâmbio na Universidade do Porto (Portugal) e

mestrado em Ciências da Computação e Matemática Computacional na própria USP. Surgiram novas amizades nessa minha nova fase, sendo as primeiras na república Toku-Tukano e depois em todos os lugares possíveis, da aula de Cálculo I ao TUSCA. Sem dúvidas, cada uma dessas pessoas contribuiu positivamente na minha formação pessoal e/ou profissional. Em particular, eu gostaria de enfatizar a importância de um dos meus primeiros mentores da área de Estatística, o professor e amigo Marinho Gomes de Andrade Filho.

Primeiro semestre de 2013, últimos meses em São Carlos e já com a expectativa de fazer o doutorado em Estatística na Universitat de València (Espanha). Durante esses meses uma pessoa especial entra na minha vida, primeiro de forma distante e discreta, mas logo insuportavelmente impossível de tê-la longe de mim. Essa pessoa é Jéssica Letícia Pavani, inicialmente uma colega que depois se converteu em uma amiga, confidente, namorada, noiva e, mais recentemente, esposa. Pitchuca, obrigado por entrar na minha vida de uma forma tão improvável e permanecer nela tão intensamente.

España

22 de agosto de 2013, esa fue la fecha de mi llegada a Valencia. Era un mundo nuevo para mi, lugares diferentes, personas de otras culturas y otros estilos de vida, nuevos desafíos y aprendizajes. Así fue como mi aventura en el doctorado comenzó y junto a ella otro máster, ahora en Bioestadística.

Los dos primeros años fueron muy largos, centrándome principalmente en el máster y sufriendo los miles de kilómetros que me separaban de mis seres queridos, en especial de mi amada. Conocí, desde el comienzo de mi nueva fase, a un personaje que muy pronto se ha convertido en mi mejor amigo, el gran matemático brasileño

Sheldon Miriel Gil Dantas. Este tío, además de ser un figura y un punto de acumulación de personas, también me hacía sentir más cerca de Brasil.

El tiempo pasó y nuevos amigos entraron en mi vida. Abel, Adina, Alex, Alfonso, Andrews, Angela, Blanca, Carles, Carme, Carol, Consuelo, Daiane, Daniel, Diana, Elena, Enric, Francesca, Guillermo, Hèctor, Iosu, Jaime, Joaquín, Juanfran, Julio, Karen, Ludmilla, Manuel, Marie-Christine, Marise, Marta, Miguel, Omar, Pedro, Rafaela, Rayd, Rufo, Sara, Thais y Walter, gracias a cada uno de vosotros por hacer mi camino doctoral más acogedor y menos estresante.

En este camino, quisera destacar a Carmen Armero y Anabel Forte, que además de ser mis directoras de tesis y jefas, se han convertido en verdaderas amigas. Siempre estuvisteis dispuestas a aconsejarme, enseñarme, escucharme y ayudarme, mi gratitud a vosotras es eterna. Aunque inconscientemente, en varias ocasiones mi profundización Bayesiana, también de la vida, fue fruto de vuestros debates durante nuestras reuniones (yo sólo observaba con una sonrisa en la mente). Por todos esos momentos y por confiar en mi trabajo, de nuevo, mil gracias.

Aprovecho la ocasión para expresar mi más profundo agradecimiento a Montse Rué. Ella apareció en mi vida a través de un trabajo de colaboración y se ha convertido en un pilar, que siempre ha tenido una palabra de apoyo y de ánimo para mi. Extiendo este agradecimiento a mis amigos de Lleida por toda la hospitalidad y también a JP que me está ayudando muchísimo en mi próximo reto.

También tengo que dar las gracias a todo el personal docente del máster en Bioestadística por tantas enseñanzas y por la capacidad de contar la estadística de una forma tan envolvente. Gracias a las

chicas de la secretaría del departamento de Estadística e Investigación Operativa por el apoyo de gestión, que nos hace la vida más fácil.

La segunda mitad de esta experiencia doctoral fue más agradable, pues mi eterna compañera y yo empezamos a vivir juntos. Tras dos años de distancia, ahora teníamos dos años de unión total. Poco a poco fuimos superando todos los obstáculos, incluso un período donde nos tuvimos que distanciar, esta vez mi destino era París. Después de todos estos contratiempos y de mucho trabajo duro, entramos en nuestra recta final, ella con el trabajo fin de máster y yo con la tesis. Letícia, mi cómplice, una vez más, gracias por entrar en mi vida de una forma tan improbable y quedarte en ella tan intensamente.

France

Tout d'abord, je voudrais remercier Adam M. Johansen (University of Warwick) et François Septier (IMT Lille Douai) pour m'indiquer le chemin que je devais suivre dans le contexte des méthodes séquentielles. Ce chemin m'a emmené à l'École Nationale de la Statistique et de l'Administration Économique. Trois mois, ce fut le temps de mon expérience française.

Ma plus profonde gratitude à Nicolas Chopin. C'était à la fois un honneur et un plaisir de travailler sous sa direction, et je lui suis infiniment reconnaissant pour sa disponibilité, son hospitalité et sa confiance.

Je remercie tous ceux qui ont rendu mon séjour à Paris (en particulier dans le bureau F14 du CREST, qui a été un "nid" pour plusieurs jeunes chercheurs Bayésienne) agréable et rentable.

Naturellement, cette expérience fut un succès total grâce au soutien inconditionnel de ma chère et bien-aimée épouse. Jéssica, merci pour être entré dans ma vie d'une manière si improbable et y être resté si intensément.

Reviewers

I want also to thank Carmen María Cadarso Suárez, Guadalupe Gómez Melis, Giovani Loiola da Silva, James McGree, José Domingo Bermúdez Edo, and Virgilio Gómez Rubio for your valuable comments. I am sure your reviews have improved this thesis.

*"If you can't explain it simply,
you don't understand it well enough"*

Albert Einstein

UNIVERSITAT DE VALÈNCIA

Abstract

Faculty of Mathematics
Department of Statistics and Operations Research

Doctor of Philosophy in Statistics and Optimization

The statistical analysis of the information generated by medical follow-up is a very important challenge in the field of personalised medicine. As the evolutionary course of a patient's disease progresses, its medical follow-up generates more and more information that should be processed immediately in order to review and update its prognosis and treatment.

Our objective in this thesis focuses on this update process through sequential inference methods for joint models of longitudinal and time-to-event data from a Bayesian perspective. More specifically, we propose the use of sequential Monte Carlo methods for static parameter joint models in order to update the posterior distribution of the parameters, hyperparameters, and random effects with the intention of reducing computation time in each update of the inferential process.

Our proposal is very general and can be easily applied to most popular joint models approaches. We illustrate our research with two different studies: (i) a joint model for longitudinal data with informative dropout simulated through an own novel mechanism, and (ii) a joint model with competing risk events for a real problem about patients receiving mechanical ventilation in intensive care units.

UNIVERSITAT DE VALÈNCIA

Resumen

Facultad de Ciencias Matemáticas
Departamento de Estadística e Investigación Operativa

Doctor en Estadística y Optimización

El análisis estadístico de la información generada por el seguimiento médico de una enfermedad es un reto muy importante en el ámbito de la medicina personalizada. A medida que avanza el curso evolutivo de la enfermedad en un paciente, su seguimiento genera cada vez más información que debe ser procesada inmediatamente para revisar y actualizar su pronóstico y tratamiento.

Nuestro objetivo en esta tesis se centra en dicho proceso de actualización a través de métodos de inferencia secuencial en modelos conjuntos de datos longitudinales y de supervivencia desde una perspectiva Bayesiana. En concreto, proponemos la utilización de métodos secuenciales de Monte Carlo adaptados a modelos conjuntos con parámetros estáticos (independientes del tiempo) para actualizar la distribución *a posteriori* de los parámetros, hiperparámetros y efectos aleatorios con la intención de reducir el tiempo de computación en cada actualización del proceso inferencial.

Nuestra propuesta es muy general y puede aplicarse de forma muy sencilla a las modelizaciones longitudinales y de supervivencia conjuntas más populares en la literatura científica del tema. Utilizamos dos estudios diferentes para ilustrar nuestra propuesta: (i) un modelo conjunto para datos longitudinales con pérdida de seguimiento informativa simulados a través de un mecanismo novedoso propio y (ii) un modelo conjunto para eventos con riesgo competitivos para un problema real sobre pacientes que reciben ventilación mecánica en unidades de cuidados intensivos.

UNIVERSITAT DE VALÈNCIA

Resum

Facultat de Ciències Matemàtiques
Departament d'Estadística i Investigació Operativa

Doctor en Estadística i Optimizació

L'anàlisi estadística de la informació generada pel seguiment mèdic és un repte molt important en l'àmbit de la medicina personalitzada. A mesura que avança el curs evolutiu de la malaltia d'un pacient, el seu seguiment mèdic genera més i més informació que caldria processar immediatament per tal de revisar i actualitzar el seu pronòstic i tractament.

El nostre objectiu en aquesta tesi se centra en aquest procés d'actualització mitjançant mètodes d'inferència seqüencial en models conjunts de dades longitudinals i de supervivència des d'una perspectiva Bayesiana. En concret, proposarem la utilització de mètodes seqüencials de Monte Carlo adaptats a models conjunts amb paràmetres estàtics (independents del temps) per tal d'actualitzar la distribució *a posteriori* dels paràmetres, hiperparàmetres i efectes aleatoris a fi de reduir el temps de computació en cada actualització del procés inferencial.

La nostra proposta és molt general i pot aplicar-se de forma molt senzilla a les modelitzacions longitudinals i de supervivència conjuntes més populars en la literatura científica del tema. Utilitzarem dos estudis diferents per il·lustrar la nostra reserca: (i) un model conjunt per a dades longitudinals amb pèrdua de seguiment informativa simulades amb un mecanisme original propi i (ii) un model conjunt per a esdeveniments de risc competitius per a un problema real sobre pacients que reben ventilació mecànica en unitats de vigilància intensiva.

Contents

List of Figures

List of Tables

Notation

Symbols and nomenclatures

$\propto$	Proportionality.
$\sim$	Has distribution.
$\in$	Is an element of.
$\lim$	Limit.
$\mathbb{Z}$	Set of integer numbers.
$\Re$	Set of real numbers.
N	Number of individuals.
i	Index for individuals.
v	Index for competitive events.
k	Index for particles.
t	Time in process.
Δt	Incremental time.
C	Censoring time.
T^*	Time until the event of interest.
T	Observed event time, $\min(T^*, C)$.
δ	Event indicator.
n	Number of longitudinal observations.
g	Number of new longitudinal observations.
t_j	jth observed longitudinal time.
y_j	Longitudinal observation at time t_j.
$t_{1:n}$	First n observed longitudinal times.
$y_{1:n}$	Longitudinal observations at times $t_{1:n}$.

$\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2$	Sets of observed data.
$\mathcal{D}_i^g$	Set of the g new observed data of the individual i.
ϵ_j	Term error at time t_j.
$\epsilon_{1:n}$	Term errors at times $t_{1:n}$.
σ^2	Longitudinal variance error term.
β's, γ's	Fixed effects parameters.
$\boldsymbol{b}, b_0, b_1$	Random effects.
σ_0^2, σ_1^2	Random effects variances term.
λ's, ν's	Parameters of the baseline hazard function.
α's	Association parameters between processes.
$\boldsymbol{\theta}$	Vector of parameters and hyperparameters.
K	Size of the marginal posterior samples of $\boldsymbol{\theta}$.
K_T	Threshold (degeneracy criterion).
M	Number of replicated particles (*resampling* step).
L	Number of integration nodes (Monte Carlo approach).
Q	Number of points (Gauss-Legendre quadrature rule).
u	Simulated nodes or points.
$\boldsymbol{\theta}^{(k)}$	Components of the kth particle.
$\widetilde{w}^{(k)}$	kth unnormalised incremental importance weight.
$w^{(k)}$	kth normalised incremental importance weight.
$f(\mathcal{D} \mid \boldsymbol{b}, \boldsymbol{\theta})$	Likelihood function of $(\boldsymbol{b}, \boldsymbol{\theta})$.
$f(\mathcal{D} \mid \boldsymbol{\theta})$	Marginal likelihood function of $\boldsymbol{\theta}$.
$\hat{f}(\mathcal{D} \mid \boldsymbol{\theta})$	Approximation of the marginal likelihood function of $\boldsymbol{\theta}$.
$\pi(\boldsymbol{\theta})$	Prior distribution of $\boldsymbol{\theta}$.
$m(\mathcal{D})$	Normalising constant.
$\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D})$	Posterior distribution of $(\boldsymbol{b}, \boldsymbol{\theta})$.
$\pi(\boldsymbol{b} \mid \mathcal{D})$	Marginal posterior distribution of $\boldsymbol{b}$.
$\pi(\boldsymbol{\theta} \mid \mathcal{D})$	Marginal posterior distribution of $\boldsymbol{\theta}$.
$\pi(\boldsymbol{b} \mid \mathcal{D}, \boldsymbol{\theta})$	Conditional posterior distribution of $\boldsymbol{b}$ given $\boldsymbol{\theta}$.
$\boldsymbol{y}$	Longitudinal process.
$\boldsymbol{s}$	Time-to-event process.
$\boldsymbol{x}$	Vector of covariates.
$f(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\theta} \mid \boldsymbol{x})$	Full joint probability distribution of $(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\theta})$ given $\boldsymbol{x}$.
$f(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x})$	Conditional joint distribution of $(\boldsymbol{y}, \boldsymbol{s})$ given $\boldsymbol{b}, \boldsymbol{\theta}$, and $\boldsymbol{x}$.
$f(\boldsymbol{b} \mid \boldsymbol{\theta}, \boldsymbol{x})$	Distribution of $\boldsymbol{b}$ given $\boldsymbol{\theta}$ and $\boldsymbol{x}$.
$f(\boldsymbol{y} \mid \cdot)$	Marginal distribution of $\boldsymbol{y}$.
$f(\boldsymbol{s} \mid \cdot)$	Marginal distribution of $\boldsymbol{s}$.
$\mu(t)$	True longitudinal value at time t.

$p(t)$	Probability for the Bernoulli process at time t.
$h(t \mid \cdot)$	Hazard function at time t.
$h_0(t)$	Baseline hazard function at time t.
$S(t \mid \cdot)$	Survival function at time t.
$F(t \mid \cdot)$	Distribution function at time t.
$H(t \mid \cdot)$	Cumulative hazard function at time t.
$P(\cdot)$	Probability.
$P(\theta > 0 \mid \mathcal{D})$	Posterior probability that parameter θ is positive.
$\mathcal{P}(\cdot)$	Generic probability distribution.

Operators and functions

$A^\top$	Transpose of matrix (or vector) A.
$F^{-1}(\cdot)$	Inverse of the function $F(\cdot)$.
$z(\cdot)$	Link function.
$\eta(\cdot)$	Linear predictor.
$\mathrm{E}(X)$	Expectation of the random variable X.
$\mathrm{Var}(X)$	Variance of the random variable X.
$\min(a, b)$	Minimum between a and b.
$\|a_{1:n}\|^2$	Euclidean distance: $\sum_{j=1}^{n} a_j^2$.
$x \to a$	x converges to a.
$I(C(x))$	Indicator function of the set C (1 if $x \in C$, 0 otherwise).
$\log(\cdot)$	Logarithmic function of base e (ln).
$\exp(\cdot)$	Exponential function.
$\Gamma(\cdot)$	Gamma function.
$\mathrm{diag}(a)$	Square diagonal matrix with a on the main diagonal.

Usual probability distributions

Bernoulli	$\mathcal{B}(p)$.
Gamma	$\mathcal{G}(\alpha, \beta)$.
Normal	$\mathcal{N}(\mu, \sigma^2)$.
Uniform	$\mathcal{U}(a, b)$.
Weibull	$\mathcal{W}(\lambda, \nu)$.

Abbreviations

AI	Asynchronies index.
CIF	Cumulative incidence function.
ESS	Effective sample size.
ICU	Intensive care units.
IBIS	Iterated batch importance sampling.
JAGS	Just Another Gibbs Sampler.
SOFA*	Logarithm of (SOFA+1).
MAR	Missing at random.
MC	Monte Carlo.
MCAR	Missing completely at random.
MCMC	Markov chain Monte Carlo.
MNAR	Missing not at random.
MV	Mechanical ventilation.
QMC	Quasi-Monte Carlo.
SD	Standard deviation.
SOFA	Sequential organ failure assessment.
SMC	Sequential Monte Carlo.
WinBUGS	Windows-based Bayesian inference using Gibbs sampling.

In memory of my grandmother Maria.

Background

The motivation of this thesis follows the current trend in medical practices towards personalised medicine[1] (Sharratt, 2015). This term is well defined according to the *Personalized Medicine Coalition*[2] as

> [Personalised medicine is] an evolving field in which physicians use diagnostic tests to determine which medical treatments will work best for each patient. By combining the data from those tests with an individual's medical history, circumstances and values, health care providers can develop targeted treatment and prevention plans.

More briefly, following the *National Academy of Sciences*[3]

---

[1]Also referred to as precision medicine, stratified medicine, individualised medicine, or P4 medicine.

[2]PMC: `http://www.personalizedmedicinecoalition.org`

[3]NAS: `http://www.nasonline.org`

> [Personalised medicine is] the use of genomic, epige-
> nomic, exposure and other data to define individual pat-
> terns of disease, potentially leading to better individual
> treatment.

Personalised medicine may be seen as the tailoring of medical treat-
ments to specific individuals as well as the needs and preferences
of a patient during all stages of care, including prevention, diagno-
sis, treatment, and follow-up (Food and Administration, 2014). In
contrast, we have the population medicine which can be understood
as the study of a group of individuals in order to extrapolate the
findings to the general population (Mega *et al.*, 2014).

The practice of personalised medicine is considered as the new era of
medicine. However, this idea is not new and has been practiced by
clinicians ever since the dawn of western medicine over 2000 years
ago (Murugan, 2015). Indeed, the famous Greek physician Hip-
pocrates of Cos (460 B.C. - 370 B.C.) is considered the precursor of
personalised medicine as well as the "father of western medicine".
At that time, he already glimpsed about the "individuality" of dis-
eases and the importance of the prescription of different medicines
to different patients (Schiefsky, 2005).

Statistical science is an essential part of medical research that has
been used in modern medicine, drug development, and epidemiol-
ogy (Chakra-Borty and Moodie, 2013; Lu *et al.*, 2015). In recent
decades, the statistical methods have greatly contributed to the
comprehension of the epidemiology of various diseases as well as

to the connection between them and biomarkers[4] and/or symptoms
(Cho *et al.*, 2012; Collette *et al.*, 2012; Zhao and Zeng, 2013; Jain,
2015). Within the diversity of methodologies that can be applied in
this field, some of them may be highlighted given their importance
in approaching personalised medicine. These are the models for lon-
gitudinal data (Bandyopadhyay *et al.*, 2011; Verbeke *et al.*, 2014) or
time-to-event data (Bewick *et al.*, 2004) or models for jointly study-
ing both type of data (De Gruttola and Tu, 1994; Tsiatis *et al.*, 1995;
Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Henderson
et al., 2000; Tsiatis and Davidian, 2004; Ye *et al.*, 2008; Rizopoulos,
2012b).

The next two sections of this chapter are designed to describe, in a
separate form, the key ideas behind longitudinal and time-to-event
data. Then, in Section 1.3 we will discuss the importance of a
joint analysis of both types of data. Finally, in Section 1.4 we will
introduce the Bayesian perspective for this joint modelling as well
as the need for sequential update methodologies.

1.1 Longitudinal data

Longitudinal data[5] are a particular type of correlated data where a
given set of variables are repeatedly measured over time in the same

---

[4]A biomarker is a characteristic that is objectively measured and evalu-
ated as an indicator of a normal biologic process, disease process, or biological
response to a therapeutic intervention. Biomarkers can be used to reduce un-
certainty and guide clinical care.

[5]Also called panel data, growth curve analysis, or multilevel analysis.

sampling unit. Longitudinal data analysis confronts with cross-sectional studies in which a single observation is measured for each individual. Longitudinal approaches can work with different correlation structures, such as serial correlation, shared random effects, transition (Markov) models, latent classes, clustering, etc. (Verbeke and Molenberghs, 2000).

One of the most important aspects of this type of data is the natural hierarchical structure of the variability with different levels of interest. For instance, we can consider as a first level effect a group mean response for all individuals over time, while a second level captures individual-specific features using, for instance, random effects (Gelman and Hill, 2006). The incorporation of individual sources of variability provides the ability to predict the trajectory of individual responses and understand how do them change with respect to the general mean in the population of interest (Laird and Ware, 1982; Verbeke *et al.*, 2001). Furthermore, there is also the possibility for describing how the mean response changes in the population of interest. All these characteristics make the longitudinal models extremely convenient in many research fields, e.g. medicine, public health, education, business, economics, psychology, and biology (Weiss, 2005).

Missing data[6] can greatly affect the longitudinal analysis depending on the processes causing them. Missing data can be classified into three types: missing completely at random (MCAR); missing at random (MAR); or missing not at random (MNAR) (Little and Rubin,

---

[6]Data that are intended to be observed, but are unobserved for some reason.

2002). The last case, MNAR, is the most important in the longitudinal studies, since it can be hiding some latent process which, if ignored, can produce biased estimates and predictions. This is why they are also known as *nonignorable missing data* (for reviews, see Molenberghs and Kenward, 2007). Time-to-event models, described in the next section, can be useful for modelling the process that generates the MNAR data, and hence avoiding the bias in the results (Ibrahim and Molenberghs, 2009).

1.2   Time-to-event data

We consider time-to-event data whenever we are interested in the time until an event of interest occurs. As pointed out by Collett (2003), this type of data arises in a great number of applied fields, such as medicine, public health, epidemiology, biology, environmental sciences, engineering, economics, actuarial sciences, management, demography, and social sciences

The nomenclature *time-to-event* analysis may change according to the area of research, e.g. it is usually referred to as *survival* analysis in medicine and biology (Box-Steffensmeier and Jones, 2004; Kleinbaum and Klein, 2012) and *reliability* analysis in engineering and industrial studies (Jewell *et al.*, 1996; Couallier *et al.*, 2013).

In numerous medical studies, survival probabilities are of primary interest, since their estimates for a specific patient can rule decision making involving specific interventions (Rizopoulos, 2011).

Commonly, the standard statistical procedures are not amenable for time-to-event data due to its particular features: the response variable is positively skewed and some observations are typically incomplete in the sense that some factor (external to the study or intentionally planned) prevents the event of interest from being observed (Lee and Wang, 2013). Some situations that illustrate those types of behaviour are: the study ends without the patient having experienced the event, an intervening event that occurs prohibiting further observation on the patient, the patient can be missing at some time point, or even the patient withdraws from the study (Merrill, 2015).

Depending on the reason why the event has not been observed we will refer to *censoring* or *truncation*. To illustrate the differences between them, we consider a toy example which focuses on children in the kindergarten, from 2 to 5 years-old, where our event of interest is the age at which these children learn to write their own name.

Let us suppose that children are annually evaluated to know if they have learnt to write their name. Evaluation for children aged 2 and 5 are also included in the analysis. In this example, we will consider only six children randomly chosen from a hypothetical population. Let us assume that child 1 moves to a different school at age 2.5 without having experienced the event of interest. In the initial examination (at 2 years-old), it was verified that child 3 already knew how to write his/her own name. Children 2, 4, and 5 experienced the event of interest within the study interval. In addition, the last evaluation (at 5 years-old) revealed that the child 6 had not yet

learned to write his/her own name. Figure 1.1 shows the data from this example.
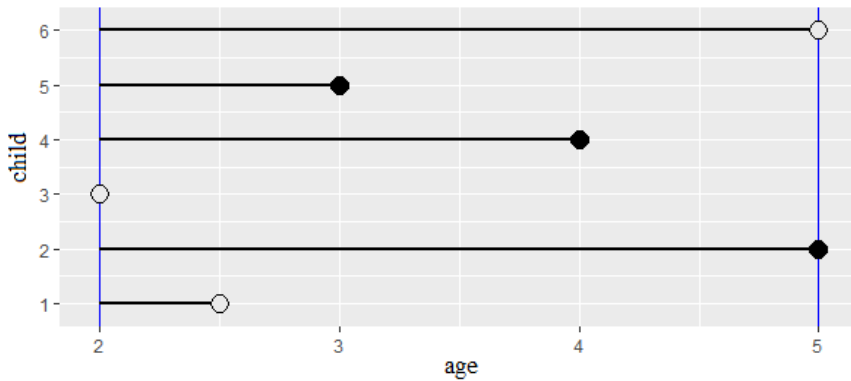
FIGURE 1.1: Age, in years, at which each child in the study learns to write their own name. Blue solid vertical lines indicate the learning period in the kindergarten. Solid circles represents observed times and open circles censored times.

We will refer to censoring data whenever the time of the event is only partially known (Leung *et al.*, 1997). This is, we know that it is larger (*right censoring*), smaller (*left censoring*), or within an interval (*interval censoring*) of observed times (Klein and Moeschberger, 2003). In our example, time for child 1 is right-censored because he/she left the study without having experienced the event of interest. Time for child 6 is also right-censored, but now because he/she has not experienced the event when the study ends. Time for child 3 is left-censored, since he/she knew how to write his/her name at the time of his/her entrance in the study. Times for children 2, 4 and 5 are interval-censored because we only know that these children had learnt to write their own name in a time between two consecutive annual evaluations.

Truncation is a variant of censoring, where the incomplete nature of the observations happens because of a systematic selection process inherent to the study design (Andersen *et al.*, 1993). In practice, truncation imposes restrictions in the limits of the period of study and only individuals with observed event time within them are considered in the analysis (Klein and Moeschberger, 2003). Hence, we can basically limit the period of study in three ways: only an upper limit (*right truncation*), only a lower limit (*left truncation*), or an interval limit (*interval truncation*).

Going back to our toy example, let us imagine that two 3-years-old children enter the study and it is verified through the learning examination that one of them already knows how to write his/her own name. In this case, the child who passed the exam has a right truncation time, since he/she experienced the event of interest before entering the study. On the other hand, we already know that the other new child is not able to write his/her own name until 3 years-old, so this child has a left truncation time (or delayed entry time). Now let us suppose that child 1 at age 4 is transferred back to our kindergarten. This child took the learning examination and was approved. In this case, child 1 has a interval truncation time, since we know he/she experienced the event of interest after 2 and before 4 years-old. It is important to remark that truncated data leads to conditional probabilities/estimations, since we know that the event time is limited superiorly and/or inferiorly by truncations.

Time-to-event analysis also allows the inclusion of covariates to improve the probabilistic modelling of the occurrence of the event of

interest. A great challenge in this sense is the incorporation of time-varying covariates, mainly those of *endogenous* (or *internal*) nature in which traditional approaches are not applicable (Molenberghs *et al.*, 2014). This is due to the fact that the occurrence of the event of interest at a time point affects (or prevents from being observed) the values of these covariates (Rizopoulos, 2012b).

Typically, biomarkers directly related to the risk of the event of interest are endogenous time-varying covariates (Kalbfleisch and Prentice, 2002). They usually have a longitudinal nature that can be incorporated into the time-to-event analysis using joint models.

1.3 Joint models for longitudinal and time-to-event data

Joint modelling has recently attracted great attention to the statistical community, especially in the area of biostatistical research (Rizopoulos *et al.*, 2010). Yu *et al.* (2004) discuss some of the main objectives of this type of models. Among them we highlight

- The study of a longitudinal response variable and its association with baseline covariates, treatments, and specific characteristics of the individuals of the population sample avoiding the possible bias due to MNAR data, which are modelled through a time-to-event process (see Section 1.1);

- Characterising the risk of an event with regard to a longitudinal variable (endogenous covariate) or to the time when a

longitudinal variable was observed (surrogate endpoint) (see Section 1.2).

In both cases, the final interest is usually to make estimation and/or prediction of longitudinal individual or population trajectories and/or relevant outcomes associated to the occurrence of the event of interest. Moreover, we could consider a general approach that takes both objectives into account.

Currently, many studies combine the power of the junction of both types of data, since a separate analysis of the longitudinal and time-to-event data may lead to inefficient or biased results (Yu *et al.*, 2008; Ibrahim *et al.*, 2010; Wu *et al.*, 2012).

Research literature for joint models uses the frequentist (or classical) and Bayesian approaches to estimate and predict information of interest (for interesting reviews up to date, see Tsiatis and Davidian, 2004; Neuhaus *et al.*, 2009). In particular, this modelling is relatively new for the Bayesian approach and so it is devoid of some more in-depth research topics, such as sensitivity analysis to the elicitation of prior distributions, sequential update, model validation, and model selection.

A key feature in real studies with this type of models is the dynamic nature in which data become available. This is clear in biomedical studies where data usually come from individual follow-ups over time. Thus, when new information of a given patient is collected, physicians are interested in updating the relevant estimated and/or predicted outcomes.

Dynamic inference is an inherent difficulty within the frequentist joint modelling, since the number and timing of interim analyses directly affect some frequentist properties (Lee and Chu, 2012). Hence, the current literature has proposed two paths: (i) taking asymptotic assumptions in order to only update the subject-specific effects (Rizopoulos, 2011; Mauguen *et al.*, 2013; Andrinopoulou *et al.*, 2015a; Barrett and Su, 2017) or (ii) moving towards a Bayesian approach (Yu *et al.*, 2008; Proust-Lima and Taylor, 2009; Rizopoulos *et al.*, 2014; Andrinopoulou *et al.*, 2015b).

In the next section, we will introduce some advantages of the Bayesian approach as well as its sequential learning for dynamic inference.

1.4   A Bayesian view of joint models

Bayesian statistics is founded upon the premise that all unknown quantities (or sources of uncertainty) should be expressed and measured by probability distributions (Bernardo and Smith, 1994). In other words, it associates probability measures to any random quantity, parameter, event, hypothesis, model, etc. (for reviews, see Loredo, 1990, 1992).

In the case of joint models for longitudinal and time-to-event data, the Bayesian perspective makes it possible to incorporate prior information to the study, thus improving and enhancing estimation and prediction of any outcome of interest (Guo and Carlin, 2004). In particular, we can also estimate and predict characteristics of the

longitudinal variable for individuals in the current sample or even for new individuals, from the same study population, that could enter to the study. However, the more relevant Bayesian issues deal with the direct estimation of the survival function, the prediction of survival times, or the computation of other posterior distributions for relevant probabilities or rates. The frequentist approach for this type of models is extremely complex because the probabilistic characteristics of the relevant estimators are practically impossible to compute (Ibrahim *et al.*, 2001).

The Bayesian approach of joint models is gaining considerable attention from the scientific community, mainly in the medical sciences (Brown and Ibrahim, 2003; Ibrahim *et al.*, 2004; Hu *et al.*, 2009; Huang *et al.*, 2010, 2011; Zhu *et al.*, 2012; Baghfalaki *et al.*, 2014; Huang *et al.*, 2014; Armero *et al.*, 2016a,b; Martins *et al.*, 2016, and references therein). One of the main factors for the proliferation of the Bayesian approach for joint models is its conceptual simplicity (Gould *et al.*, 2014). Moreover, the improvement of Bayesian computational methods, the increase of the processing capacity, and the development of Bayesian statistical software and packages has contributed to the use of multilevel[7] models, which can be considered as a natural framework for joint models.

Another motivation in favour of the Bayesian methodology is our interest in dynamic inference. The reason is simple and conceptual. The *posterior distribution* is the most relevant element of the

---

[7]Other related names: hierarchical linear models, nested models, random parameter models, random coefficient, random-effects models, mixed models, or split-plot designs.

Bayesian approach, because besides being the probabilistic representation of the total knowledge about the parameters after the data and any other relevant information have been considered (Glickman and van Dyk, 2007), it is also the starting point of all relevant inferences. The posterior distribution combines the *likelihood function*, interpreted as the information about the parameters contained in the data, and the *prior distribution*, which represents prior expert knowledge about the parameters. Bayesian learning consists of the sequential application of Bayes' theorem for updating the posterior distribution with information provided by new experimental data (Barber, 2012).

Nowadays, most of the packages for joint models[8] are implemented from a frequentist perspective and do not incorporate mechanisms of dynamic inference. The only package that uses a Bayesian perspective is `JMbayes`, which, despite it does not make fully Bayesian dynamic inference, proposes a mechanism based on asymptotic theory for dynamic prediction (for more details, see Rizopoulos, 2011). Bayesian inference within joint modelling can also be done using general Bayesian software/packages, such as the BUGS language (`Win`/`OpenBUGS` (Lunn *et al.*, 2000) and `JAGS` (Plummer, 2003)), `Stan` (Hoffman and Gelman, 2014), and `INLA` (Rue *et al.*, 2009), but so far none of them incorporates dynamical procedures.

---

[8]In `R`: `lcmm` (Proust-Lima *et al.*, 2010), `JM` (Rizopoulos, 2010), `joineR` (Philipson *et al.*, 2012), `JMbayes` (Rizopoulos, 2012a), `JMdesign` (Cornea *et al.*, 2014), `JSM` (Xu *et al.*, 2016), `joint.Cox` (Emura, 2016), and `frailtypack` (Rondeau *et al.*, 2017). In `SAS`: `JMFit` macro (Zhang *et al.*, 2016). In `STATA`: `stjm` module (Crowther *et al.*, 2013).

Our main objective in this thesis is to propose an entirely inferential and predictive Bayesian analysis for joint models by means of a dynamic update methodology based on sequential Monte Carlo[9] methods (Cappé *et al.*, 2007). Therefore, our challenge is to combine the flexibility of the Bayesian approach for joint models of longitudinal and time-to-event data with a learning procedure constructed through sequential methods.

As far as we know this thesis is the first proposal that fully integrates the Bayesian joint models with sequential Monte Carlo methods.

1.5   Outline

After this introductory chapter aiming to briefly introduce Bayesian joint modelling for longitudinal and time-to-event data and motivate the need of sequential Monte Carlo methods to achieve dynamic estimation and prediction in such models, the rest of this thesis is structured as follows. Chapter 2 reviews the Bayesian paradigm and presents sequential Monte Carlo methods. Chapter 3 contains the special features of sequential Monte Carlo methods tailored to the framework of joint modelling for longitudinal and time-to-event data. This chapter may be regarded as the core of the thesis. Chapters 4 and 5 illustrate our proposals. In particular, Chapter 4 provides the development of a simulation mechanism for generate data from a joint model with longitudinal objective and nonignorable

---

[9]Other related names: particle filtering, Monte Carlo filter, survival of the fittest, sequential imputations, condensation, bootstrap filter, or sequential importance resampling.

dropout. From a generated data set with this mechanism, we exemplify the benefits of a joint analysis for longitudinal data with informative dropout. Still in this chapter, we explore in detail the use of our proposal of sequential updating for a shared-parameter joint model. Chapter 5 explores the dynamic posterior estimation and prediction for a random-effects joint model constructed in terms of a longitudinal linear mixed submodel and a survival competing risks submodel. This is a joint model for a real study devoted to the analysis of the association between a severity marker and the events *alive discharge* and *death* for patients receiving mechanical ventilation in intensive care units. Chapter 6 presents the main conclusions and contributions of the thesis as well as a brief discussion about future research on the subject. We have also included three appendices. Appendix A contains the `JAGS` codes of the simulated model implemented in Chapters 4 and 5. Appendix B presents a list of the common probability distributions and their basic properties. Appendix C provides a posterior summary of the results obtained in simulated scenarios based on different parameter configurations. This thesis ends with a final section devoted to Bibliography.

Sequential learning

Learning theory is broadly a framework for machine learning[1] that comes from the fields of statistics and functional analysis (Mohri *et al.*, 2012). In essence, statistical learning refers to a set of approaches for inferring and predicting from available data (Hastie *et al.*, 2009; James *et al.*, 2013). In addition, it also plays an important role for sequential settings, such as classification problems (Syed *et al.*, 2009).

In particular, sequential learning can be defined in statistical terms as the mechanism to improve estimation and prediction after observing new data (Dietterich, 2002). The way of processing knowledge in the sequential learning framework is similar to the human ability to learn from previous experiences (Clegg *et al.*, 1998).

---

[1]Machine learning is a set of rules and procedures, which allows computers to act and make decisions based on data rather than being explicitly programmed to perform a certain task.

In the context of statistical inference, Figure 2.1-(a) shows a generic scheme of a non-sequential inferential process from a general model.
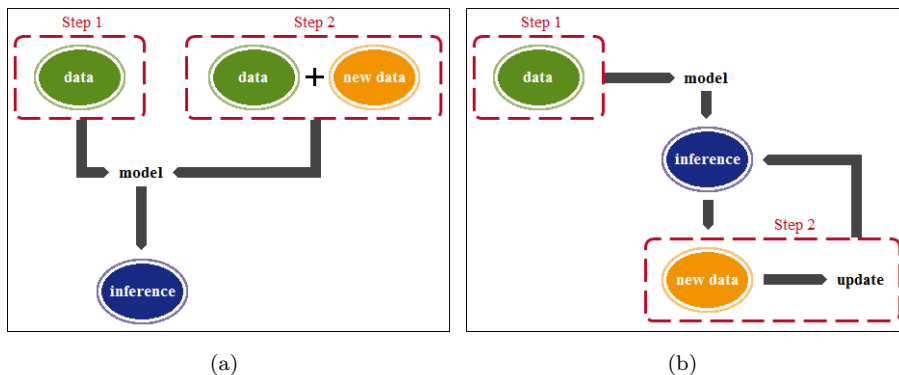
FIGURE 2.1: Non-sequential (a) and sequential (b) procedures.

In this illustration, *Step 1* works with data and *Step 2* with the previous data plus new data. Note that in *Step 2* we proceed as if all the available data were new, despite part of them were already used in *Step 1*. Figure 2.1-(b), instead, uses only new data to update the inferential process, without the necessity of a "complete oblivion" of the results previously obtained. The idea of sequential learning is to avoid the rework illustrated in Figure 2.1-(a) maintaining the accuracy of the results.

As pointed out in the introductory chapter, our objective is to develop a full Bayesian learning process for making dynamic inference and prediction in joint models for longitudinal and time-to-event data. Hence, in the following sections we will introduce the basis of the Bayesian approach as well as Bayesian sequential methods that will be adapted to the context of the joint models encompassed in this thesis.

2.1 Bayesian approach

As briefly discussed in Section 1.4, Bayesian inference contains two key ingredients. The first element is the *likelihood function* of the parametric vector $\boldsymbol{\theta}$, $f(\mathcal{D} \mid \boldsymbol{\theta})$, which summarises the information available in the data $\mathcal{D}$ about $\boldsymbol{\theta}$. The second ingredient $\pi(\boldsymbol{\theta})$, called *prior distribution* of $\boldsymbol{\theta}$, is a probability distribution that contains all the available prior expert knowledge about $\boldsymbol{\theta}$. From these two terms, the inferential process should naturally be summarised by the probability distribution of $\boldsymbol{\theta}$ after observing the value of $\mathcal{D}$. This distribution, $\pi(\boldsymbol{\theta} \mid \mathcal{D})$, is known as the *posterior distribution* of $\boldsymbol{\theta}$ and it is obtained according to the Bayes' rule:

$$\pi(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{f(\mathcal{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(\mathcal{D})} \propto f(\mathcal{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}), \qquad (2.1)$$

where $m(\mathcal{D}) = \int f(\mathcal{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$ is the normalising constant, also called *model evidence* of the data $\mathcal{D}$ (Robert, 2007). This constant that makes the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathcal{D})$ integrate to one.

The basis of the Bayesian methodology is simple and intuitive. It combines different sources of information that can be associated to the way in which the human brain works, always learning from past experiences. The Bayesian approach offers comprehensive and powerful tools to complex model estimation and prediction as well as all relevant quantities of interest derived from the inferential process (Carlin *et al.*, 2001; Gilks *et al.*, 1993).

One of the main advantages of the Bayesian methodology is that, independently of the complexity of the model, it always follows the

same structure based on the posterior distribution of $\boldsymbol{\theta}$ in (2.1). In addition, the prior distribution of $\boldsymbol{\theta}$ may incorporate prior knowledge, often based on expert or information accumulated from previous studies, which in many cases makes posterior estimates more accurate (Bayarri and Berger, 2004). Furthermore, Bayesian perspective does not need to assume asymptotic assumptions, as is common in the frequentist paradigm (Ibrahim *et al.*, 2001).

Even with all these advantages, the beginning of the Bayesian methodology was frustrating because, despite being theoretically appealing, it was almost always inapplicable in practice. The big challenge was (and in some sense still is) to handle with nonstandard probability densities, especially in high-dimensional problems (Robert, 2014). In particular, the main difficulty was to calculate the normalising constant $m(\mathcal{D})$ and to obtain a sample of the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D})$ (Robert and Casella, 2004).

After a long period of dormancy (McGrayne, 2011), the Bayesian perspective resurfaced in the early 90s due, mainly, to the evolution of technology (computers with more processing capacity and more affordable prices) and the development of stochastic integration methodology, especially Markov chain Monte Carlo (MCMC) approaches, such as Gibbs sampling and Metropolis-Hastings algorithms (Geman and Geman, 1984; Gelfand and Smith, 1990; Carlin and Chib, 1995; Gelfand, 2000; Gamerman and Lopes, 2006). These advances entailed a substantial increase of the number of publications involving Bayesian methods in many scientific areas (Robert and Casella, 2011). More recently, other Bayesian procedures based

on analytical approximations of the posterior distribution have also appeared, such as variational Bayesian methods (Beal, 2003) and integrated nested Laplace approximations (Rue *et al.*, 2009). However, for the inference in many contexts, such as most part of join models, we are still restricted to the use of MCMC methods.

Sequential inference is one of the more important scenarios where Bayesian methodology has gained very much popularity (Creal, 2012). The reason is that the Bayesian inference provides a natural, elegant, and unified approach to sequential learning (Freitas *et al.*, 1999).

For an illustration of Bayesian reasoning in sequential learning, suppose a given generic model with parametric vector $\boldsymbol{\theta}$ for which we want to make inference from only a set of available observations $\mathcal{D}_1$. The posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$, is computed by applying the Bayes' rule from the likelihood function of $\boldsymbol{\theta}$, $f(\mathcal{D}_1 \mid \boldsymbol{\theta})$, and the prior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$, as shown in (2.1). Then, in a second step, we observe a new set of observations $\mathcal{D}_2$ and we want to update our knowledge about $\boldsymbol{\theta}$ obtaining a "new" posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$. This sequential procedure can be summarised by equations (2.2) and (2.3) as follows:

$$\textit{Step 1:} \qquad \pi(\boldsymbol{\theta} \mid \mathcal{D}_1) \propto f(\mathcal{D}_1 \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{2.2}$$

$$\textit{Step 2:} \qquad \pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2) \propto f(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathcal{D}_1). \tag{2.3}$$

However, it is quite common that the "first" posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$, has not an analytical expression, i.e., we only have an approximated sample of it thus making the sequential inferential

process "not-so-easy" in practice. In such cases, equivalent inferences can be obtained through a Bayesian inferential process based on the set that integrates the old and new data:

$$\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2) \propto f(\mathcal{D}_1, \mathcal{D}_2 \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \qquad (2.4)$$

Nevertheless, this procedure is not always a real alternative because it may be computationally very costly in terms of both, time and resources.

To circumvent the problem that $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ is analytically intractable, a great number of sophisticated techniques have been proposed in recent years, in particular sequential Monte Carlo methods. These are a general class of numerical methods, which provide samples from a target distribution (let's say the posterior distribution) based on weights calculated from importance sampling and resampling mechanisms (Kantas *et al.*, 2009; Gao and Zhang, 2012).

In the next section, we will introduce sequential Monte Carlo methods and their main peculiarities.

2.2   Sequential Monte Carlo methods

Among the numerous sequential Bayesian learning approaches, the most efficient methods for inference are the sequential Monte Carlo (SMC) methods (Lopes and Tsay, 2011). SMC methods are a set of simulation-based procedures which provide an appropriate and

clever approach to sequentially update complex posterior distributions. These methods are flexible, applicable to very general settings, and their implementations are intuitive and allow parallel processing (Doucet *et al.*, 2001).

In general, SMC methods are employed in a plethora of applications involving artificial intelligence, bioinformatics, computational physics, computational science, economics and mathematical finance, engineering and robotics, machine learning, molecular chemistry, pharmacokinetic, phylogenetics, signal and image processing, simultaneous localization and mapping, target tracking, among other fields (Jouin *et al.*, 2016). In all these frameworks, the parameters and/or states[2] of interest are commonly associated with time or some similar dependent structure, more specifically known as state-space (or hidden Markov) models (Crisan and Rozovskii, 2011).

SMC methods approximate the target distribution using a set of simulated samples (particles) and their respective weights. They adopt a sequential strategy for updating that distribution incorporating the information provided by new data (Bonawitz *et al.*, 2014). The performance of these methods depends on the number of particles, in which a very large number indicates a better representation of the target distribution.

Figure 2.2 illustrates the step by step of the general SMC scheme with 12 particles (circles). In order to facilitate the understanding

---

[2]Unobserved (or hidden) process that connects the dynamic nature of the response variable with a conditional model for the observed process given the state(s).
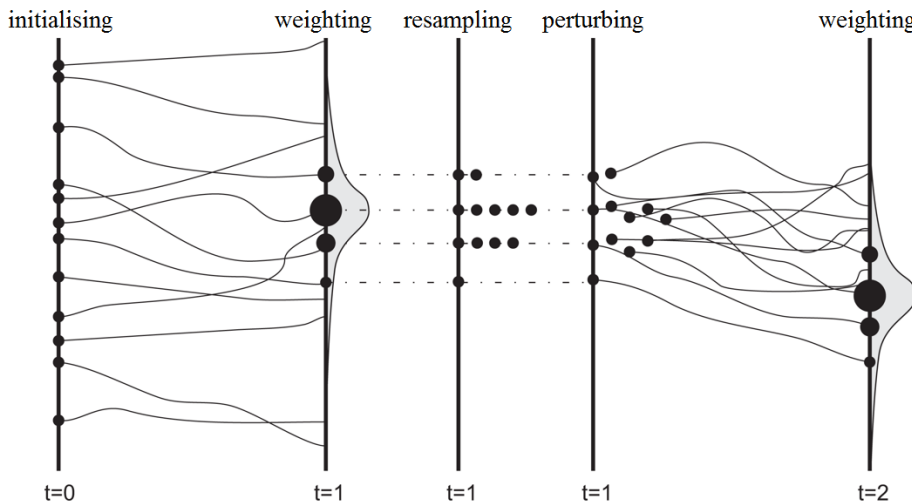
FIGURE 2.2: Sequential Monte Carlo scheme.
(Source: Montzka *et al.*, 2012)

of the graphic, let us imagine that these particles synthesise the distribution of a parameter θ that needs to be updated. The size of each circle in a given step is related to the weight of the specific particle. At $t = 0$ the particles are distributed over an interval (vertical axis) that can be interpreted as the domain of the values where the parameter θ is defined, with all its particles having equal weights. At $t = 1$ we have the first observation(s) and the sequential update procedure is initialised. The first step at $t = 1$ (*weighting*) consists in obtaining the weights of the particles according to the information provided by the new observations. In the second step at $t = 1$ (*resampling*), the set of particles is resampled with probabilities proportional to their weights which are then reset to be equally likely. In the third step at $t = 1$ (*perturbing*[3]) the particles are

---

[3] *Perturbing* step does not exist in some sequential methods, e.g. sequential importance resampling (Doucet and Johansen, 2011).

moved to avoid their accumulation in a few values. The next time point $t = 2$ represents access to new observed data and the restarting of the update procedure from the *weighting* step. Although it does not appear explicitly in Figure 2.2, before performing the *resampling* and *perturbing* steps, the "good quality" (efficiency) of the new particles should be checked (Arulampalam *et al.*, 2002). This is done according to some degeneracy criterion. The problem of degeneracy occurs when only a few particles representing the target distribution have significant weights. Further down we will display the most standard way of measuring this degeneracy.

As previously mentioned, this class of sequential methods was (and in some sense still is) primarily developed for state-space models, where parameters/states are time dependent. This is not the situation in joint models[4] where all parameters and/or hyperparameters are static in the sense that they do not change in time. For this reason we focus on SMC methods for models of static parameters (also known as static models) and we are thus faced to the proposal by Chopin (2002) and other works with the same background (Ridgeway and Madigan, 2003; Balakrishnan and Madigan, 2006; Del Moral *et al.*, 2006; Cappé *et al.*, 2008; Schäfer and Chopin, 2013; Fearnhead and Taylor, 2013; Chopin *et al.*, 2013). The algorithm proposed in Chopin (2002), called *iterated batch importance sampling* (IBIS), aims to approximate a target distribution of static parameters and hyperparameters, e.g. $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$ in (2.3), using

---

[4]Technically, we could employ non-static joint models, but this thesis only focuses on joint models with static parameters.

a sample of a "prior distribution", e.g. $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ in (2.2), through the scheme presented in Figure 2.2.

Our starting point is an approximate random sample of size K from the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$, obtained through a numerical Bayesian procedure[5]. After this initial stage and with new data, $\mathcal{D}_2$, available, the update process using the IBIS algorithm has four steps: *initialising*, *weighting*, *resampling*, and *moving* (or *perturbing*) (see Figure 2.2).

The first step (*initialising*) is used to generate a particle system $\left(\boldsymbol{\theta}^{(k)}, w^{(k)}\right)$, where $\boldsymbol{\theta}^{(k)}$ is drawn from the "first" posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$, and the weight $w^{(k)}$ of each sampled particle $\boldsymbol{\theta}^{(k)}$ is equal to $1/K$, for $k = 1, \ldots, K$. This first step is performed only once, while the next ones are iterative as more data becomes available.

The *weighting* step is based on importance sampling and resampling techniques (Rubin, 1987, 1988). Basically, they rely on an importance distribution to calculate the "changes" in the target distribution, e.g. $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$, through (incremental importance) weights. The choice of an appropriate importance distribution is the key for an efficient update to be sequentially performed. In short, it must be a good approximation of the target distribution. So provided that new data $\mathcal{D}_2$ should not alter much the inference about the parameters obtained with $\mathcal{D}_1$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ and $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$ are likely to be similar. Hence, we define the importance distribution as $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ and then update the unnormalised weights $\widetilde{w}^{(k)}$, for $k = 1, \ldots, K$,

---

[5]In this thesis, we always use the MCMC approach as the starting point.

based on the likelihood function of the new data:

$$\widetilde{w}^{(k)} \propto \frac{\pi\left(\boldsymbol{\theta}^{(k)} \mid \mathcal{D}_1, \mathcal{D}_2\right)}{\pi\left(\boldsymbol{\theta}^{(k)} \mid \mathcal{D}_1\right)} \propto \frac{f\left(\mathcal{D}_1, \mathcal{D}_2 \mid \boldsymbol{\theta}^{(k)}\right) \pi\left(\boldsymbol{\theta}^{(k)}\right)}{f\left(\mathcal{D}_1 \mid \boldsymbol{\theta}^{(k)}\right) \pi\left(\boldsymbol{\theta}^{(k)}\right)}$$

$$= \frac{f\left(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta}^{(k)}\right) f\left(\mathcal{D}_1 \mid \boldsymbol{\theta}^{(k)}\right)}{f\left(\mathcal{D}_1 \mid \boldsymbol{\theta}^{(k)}\right)} = f\left(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta}^{(k)}\right). \quad (2.5)$$

Then the weights $\widetilde{w}^{(k)}$, $k = 1, \ldots, K$, are normalised to sum to one.

Next, we have a *resampling* (or *selection*) step. The principle of resampling is simple: the particles with low normalised importance weights are discarded with a high probability, while those that remain are replicated. Figure 2.3 illustrates a generic resampling scheme.
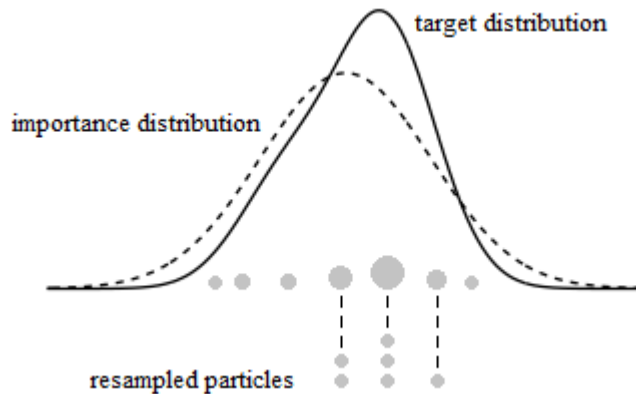
FIGURE 2.3: Graphical illustration of the *resampling* step.
(Source: modified from Cappé *et al.*, 2005)

Initially (top of the Figure 2.3), we have the target posterior distribution $\pi\left(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2\right)$ (solid line), the importance distribution $\pi\left(\boldsymbol{\theta} \mid \mathcal{D}_1\right)$ (dashed line), the particles (circles), and the weights (size of the circle) as in *weighting* step. Then (bottom of the Figure 2.3),

the particles are resampled taking into account their normalised importance weights and M particles are selected with weights reset to $1/M$ (in the Figure 2.3, $K = 7$ and $M = 6$).

The *resampling* step maintains (the majority of) the particles in regions of high probability mass then culminating with the reduction of the number of particles to represent the target distribution, and diminishing consequently the computational effort (Doucet *et al.*, 2000; Del Moral *et al.*, 2012).

It is important to remark that the *resampling* step increases the Monte Carlo variance (Chopin, 2004), but it does not change the expected value of the target distribution, i.e., the estimators are kept unbiased (Petris *et al.*, 2009). In addition, resampling eliminates the accumulation of errors over time and provides better stability for predictive distributions (Douc *et al.*, 2014). These concepts are precisely characterised by existing convergence results (Doucet and Johansen, 2011).

Many *resampling* procedures that keep under control the increasing Monte Carlo variance while preserving the unbiasedness property have been proposed (Douc and Cappé, 2005; Hol *et al.*, 2006). The most popular of them are multinomial resampling (Gordon *et al.*, 1993), residual (or remainder) resampling (Whitley, 1994; Liu and Chen, 1998), stratified resampling (Kitagawa, 1996), and systematic resampling (Whitley, 1994; Carpenter *et al.*, 1999). Since this thesis does not focus on quality and/or computational complexity of resampling methods, we will only use the multinomial resampling[6], which is the simplest approach.

Intuitively, replicating particles with large importance weights implies progressive *sample impoverishment*[7] and high correlation between the resampled particles. To avoid this problem without losing the accuracy of the results a particle rejuvenation scheme should be added (Gilks and Berzuini, 2001).

The rejuvenation of particles is achieved by means of a *perturbing* step (from now on we will refer to this step as *moving* in order to maintain the same nomenclature used in Chopin (2002)). This step strongly reduces the degeneration of the importance weights over time. It is usually performed from an MCMC kernel with posterior distribution of $\boldsymbol{\theta}$, $\pi\left(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2\right)$, as its stationary distribution (Chopin *et al.*, 2013). The key idea is to slightly move the particles in order to maintain the diversity of the samples in the parameter space. We employ an independent Metropolis-Hastings kernel in which the proposed particles are independently generated from a normal proposal (also known as instrumental or jumping) distribution of $\boldsymbol{\theta}$ based on the mean and the variance of the particles according to their weights. Other options would be to use the random-walk Metropolis-Hastings or Metropolis-within-Gibbs algorithms (Tierney, 1994; Chib and Greenberg, 1995). This second approach is advisable when some full conditional distributions are known and easy to simulate.

---

[6]Sample M "new" particles with replacement from the set of particles $\left(\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}\right)$, where the probabilities of selection are defined by $P\big(\boldsymbol{\check{\theta}}^{(r)} = \boldsymbol{\theta}^{(k)}\big) = w^{(k)}$ for $k = 1, \ldots, K$ and $r = 1, \ldots, M$. Usually K is larger than M.

[7]Also known as *weight degeneracy*. This phenomenon occurs when a few particles have large normalised weights while all remaining have negligible weights. It causes a deterioration in the Monte Carlo approximation.

Finally, the particle system $\left(\boldsymbol{\theta}^{(k)}, w^{(k)}\right)$, for $k = 1, \ldots, K$, is replaced by the new particles and their respective weights $\left(\check{\boldsymbol{\theta}}^{(r)}, w^{(r)}\right)$, for $r = 1, \ldots, M$, and the update process is finished.

An important adaptation that Chopin (2002) incorporated into the proposal by Gilks and Berzuini (2001) is that the *resampling* and *moving* steps are unnecessary when the weights of the particles have a small variance. This strategy significantly reduces the computational time of the sequential update procedure. In other words, the "first" posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$, is maintained as a good approximation for the second one, $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$. However, most of the times after initialising the update process, the sample impoverishment occurs. Hence, it is essential to establish some criterion to decide whether to continue with the same particles or perform *resampling* and *moving* steps. In practice, the (empirical) variability of the weights is usually evaluated through the *effective sample size* (also referenced as degeneracy criterion), which is defined as:

$$ESS = \left[\sum_{k=1}^{K} \left(w^{(k)} \right)^2 \right]^{-1} = \frac{\left(\sum_{k=1}^{K} \widetilde{w}^{(k)} \right)^2}{\sum_{k=1}^{K} \left(\widetilde{w}^{(k)} \right)^2}, \qquad (2.6)$$

and varies between 1 (when a normalised weight is equal to one) and K (when all the weights are equal). Thus, when ESS falls below a threshold K_T (typically $K_T = K/2$), the *resampling* and *moving* steps are triggered. Algorithm 1 shows a brief description of the SMC procedure presented in this section.

Note that the normalised weights $w^{(k)}$, for $k = 1, \ldots, K$, do not appear in the calculation of incremental weight (2.5), but we have

---

Algorithm 1: Iterated batch importance sampling

---

1 INITIALISING: draw $\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ and set $w^{(k)} \leftarrow 1/K$, $k = 1, \dots, K$.

2 WEIGHTING: from new data $\mathcal{D}_2$, calculate

$$\widetilde{w}^{(k)} \leftarrow f\left(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta}^{(k)}\right) w^{(k)},$$

and normalise the weights $w^{(k)} \leftarrow \dfrac{\widetilde{w}^{(k)}}{\sum_{l=1}^{K} \widetilde{w}^{(l)}}, \quad k = 1, \dots, K$.

if $(ESS < K_T)$ **then**

3   RESAMPLING: draw $\left(\widetilde{\boldsymbol{\theta}}^{(1)}, \dots, \widetilde{\boldsymbol{\theta}}^{(M)}\right)$ from $\left(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)}\right)$ with probabilities proportional to the normalised weights ($M \leq K$). Update $w^{(r)} \leftarrow 1/M, \ r = 1, \dots, M$.

4   MOVE: draw $\check{\boldsymbol{\theta}}^{(r)}$ from a Metropolis-Hastings kernel of invariant distribution $\pi\left(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2\right), \ r = 1, \dots, M$. Update $\boldsymbol{\theta}^{(r)} \leftarrow \check{\boldsymbol{\theta}}^{(r)}, \ r = 1, \dots, M$ and $K \leftarrow M$.

end

If new data available, return to WEIGHTING step.

---

added them in the *weighting* step (see Algorithm 1). Recall that from the second sequential update the *initialising* step is no longer activated (see Figure 2.2) and so the incorporation of these weights is important, since they allow to accumulate the information from included observations in the previous update when the *resampling* and *moving* steps were not required.

An important feature of the sequential Monte Carlo approach is that it sequentially can provide an estimate of the marginalisation constant (see (2.1)) with very little additional computation. This result is essential in model selection based on Bayes factors (Kass and Raftery, 1995). This is a very relevant and challenging issue in Statistics which is beyond the objective of this thesis. In any case,

we will briefly discuss in the next chapter the general procedure to calculate the approximate update of marginalisation constants in joint models.

Although the use of IBIS algorithm is computationally appealing by preventing us from the calculation of the posterior distribution of $\boldsymbol{\theta}$ from scratch (considering $\mathcal{D}_1$ as initial data and $\mathcal{D}_2$ as new data), it has a major limitation in our joint models framework. This is related to the use of the so-called *random effects*. A statistical artifact that, as mentioned in Section 1.1, allow to consider individual divergences from the population behaviour. Next chapter is devoted to the introduction of some key adaptations in the IBIS approach for dealing with joint models for longitudinal and time-to-event data containing subject-specific effects.

Sequential methods for Bayesian joint models

As introduced in Chapter 1, our modelling of interest is Bayesian joint models for longitudinal and time-to-event data. The major goal of these models is the connection between the longitudinal and the time-to-event processes.

There are different types of associations between both processes (Daniels and Hogan, 2008; Fitzmaurice *et al.*, 2008). Most of them can be modelled by some structure of unobserved latent variables and/or parameters (Sousa, 2011).

We start this chapter by introducing the main connection structures between longitudinal and time-to-event processes from a Bayesian perspective. Next, as the core of the thesis, we will extend the sequential procedures discussed in Section 2.2 to those types of joint models.

3.1   Bayesian joint models

Bayesian joint models for longitudinal and time-to-event data assume a full joint probability distribution

$$f\left(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\theta} \mid \boldsymbol{x}\right) = f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right) f\left(\boldsymbol{b} \mid \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\right), \qquad (3.1)$$

where $\boldsymbol{y}$ and $\boldsymbol{s}$ represent the longitudinal and the time-to-event process respectively. Random effects are denoted by $\boldsymbol{b}$, $\boldsymbol{\theta}$ represent the parameters and hyperparameters, and $\boldsymbol{x}$ is a set of covariates. Notice that each covariate in $\boldsymbol{x}$ can be related only with the longitudinal or the time-to-event process or with both of them. In (3.1), $f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right)$ is the conditional joint distribution for the processes $\boldsymbol{y}$ and $\boldsymbol{s}$ given the random effects, parameters and hyperparameters, and covariates, $f\left(\boldsymbol{b} \mid \boldsymbol{\theta}\right)$ is the conditional distribution of the random effects given $\boldsymbol{\theta}$, and $\pi\left(\boldsymbol{\theta}\right)$ the prior distribution of the parameters and the hyperparameters of the model. The conditional joint probability distribution $f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right)$ usually depend on the assumptions about the association of both processes. The different approaches that conditionally connect the longitudinal and survival processes that we will consider in this thesis are conditional, shared-parameter, and random-effects models.

In *conditional models*, the conditional joint probability distribution $f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right)$ in (3.1) is decomposed into the product of conditional and marginal distributions between longitudinal and time-to-event processes (Little, 2008). More specifically, this decomposition can be written in two opposing ways, called pattern-mixture and

selection models. Pattern-mixture models factorises the conditional joint distribution $f(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x})$ into the product of the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{s}$, $\boldsymbol{\theta}$, and $\boldsymbol{x}$, and the marginal distribution of $\boldsymbol{s}$ given $\boldsymbol{b}$, $\boldsymbol{\theta}$, and $\boldsymbol{x}$.

$$f(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}) = f(\boldsymbol{y} \mid \boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{x}) \, f(\boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}). \qquad (3.2)$$

It is important to note that the random effects are only directly connected to the time-to-event process. Figure 3.1 illustrates the general idea for this approach.
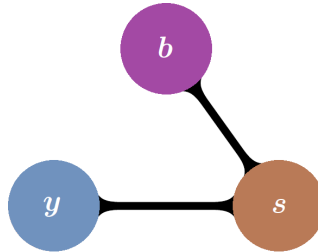
FIGURE 3.1: Relationship between the longitudinal process $\boldsymbol{y}$, time-to-event process $\boldsymbol{s}$, and random effects $\boldsymbol{b}$ using pattern-mixture models.

In general, this approach is used when the objective of the study is the longitudinal process $\boldsymbol{y}$.

On the other hand, selection models assume the decomposition of the conditional joint distribution $f(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x})$ in terms of the product of the conditional distribution of $\boldsymbol{s}$ given $\boldsymbol{y}$, $\boldsymbol{\theta}$, and $\boldsymbol{x}$, and the marginal distribution of $\boldsymbol{y}$ given $\boldsymbol{b}$, $\boldsymbol{\theta}$, and $\boldsymbol{x}$.

$$f(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}) = f(\boldsymbol{s} \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{x}) \, f(\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}). \qquad (3.3)$$

In these models, the random effects are directly connected to the longitudinal process. Figure 3.2 depicts the relationship between the components of the joint process from the selection approach.
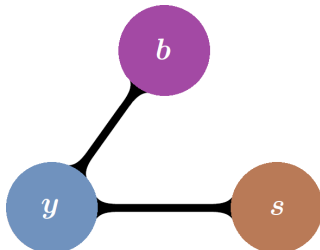
FIGURE 3.2: Relationship between the longitudinal process $\boldsymbol{y}$, time-to-event process $\boldsymbol{s}$, and random effects $\boldsymbol{b}$ using selection models.

In contrast to the pattern-mixture approach, these models usually have a time-to-event objective.

Shared-parameter models are the most popular approach that connect the longitudinal and the time-to-event processes (Wu and Carrol, 1988; Wu and Bailey, 1988; Hogan and Laird, 1997a,b, 1998; Vonesh *et al.*, 2006). In this case, both processes are considered as conditionally independent given the random effects $\boldsymbol{b}$, the parameters and hyperparameters $\boldsymbol{\theta}$, and the covariates $\boldsymbol{x}$.

$$f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right) = f\left(\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right) f\left(\boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right). \qquad (3.4)$$

Figure 3.3 represents this structure of association.

The interpretation for this approach is based on the belief that both processes are governed by a common set of underlying latent individual characteristics.
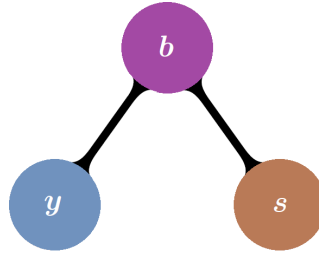
FIGURE 3.3: Relationship between the longitudinal process $\boldsymbol{y}$, time-to-event process $\boldsymbol{s}$, and random effects $\boldsymbol{b}$ using shared-parameter models.

Random-effects models are characterised by postulating conditional independence between longitudinal and time-to-event processes given the random effects of each process $\boldsymbol{b} = \left(\boldsymbol{b}_{(\boldsymbol{y})}, \boldsymbol{b}_{(\boldsymbol{s})}\right)^{\top}$, the parameters and hyperparameters $\boldsymbol{\theta}$, and the covariates $\boldsymbol{x}$, and assuming some structure of correlation between $\boldsymbol{b}_{(\boldsymbol{y})}$ and $\boldsymbol{b}_{(\boldsymbol{s})}$ (Henderson *et al.*, 2000).

$$f\left(\boldsymbol{y}, \boldsymbol{s} \mid \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{x}\right) = f\left(\boldsymbol{y} \mid \boldsymbol{b}_{(\boldsymbol{y})}, \boldsymbol{\theta}, \boldsymbol{x}\right) f\left(\boldsymbol{s} \mid \boldsymbol{b}_{(\boldsymbol{s})}, \boldsymbol{\theta}, \boldsymbol{x}\right), \qquad (3.5)$$

where the structure of correlation between $\boldsymbol{b}_{(\boldsymbol{y})}$ and $\boldsymbol{b}_{(\boldsymbol{s})}$ will be defined by means of the joint probability distribution $f\left(\boldsymbol{b} \mid \boldsymbol{\theta}\right)$. Figure 3.4 shows the relationship between processes and random effects.
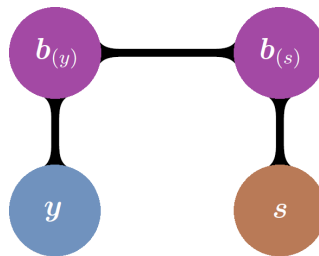
FIGURE 3.4: Relationship between the longitudinal process $\boldsymbol{y}$, time-to-event process $\boldsymbol{s}$, and random effects $\boldsymbol{b} = \left(\boldsymbol{b}_{(\boldsymbol{y})}, \boldsymbol{b}_{(\boldsymbol{s})}\right)^{\top}$ using random-effects models.

In all these approaches, the parametric vector $\boldsymbol{\theta}$ can also be particularised as $\boldsymbol{\theta}_{(\boldsymbol{y})}$ (longitudinal) and $\boldsymbol{\theta}_{(\boldsymbol{s})}$ (time-to-event).

The sequential methodology that will be presented in the next section is applicable to any of these joint modelling of association between the longitudinal and the time-to-event processes. Still, there are at least two classes of joint models that we will not be addressed in this thesis, the so-called joint latent class models (Proust-Lima *et al.*, 2009) and joint models using Gaussian copula (Ganjali and Baghfalaki, 2015). The reason is simply that these models have a differentiated structure for relating the longitudinal and time-to-event processes and so our sequential proposal would have to incorporate some additional features.

3.2   Updating the posterior information

The posterior distribution of $\boldsymbol{\theta}$ in (2.1) applied to the case of joint models can be given by the hierarchical modelling (Diggle *et al.*, 2002):

$$\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D}) \propto f(\mathcal{D} \mid \boldsymbol{b}, \boldsymbol{\theta}) f(\boldsymbol{b} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}), \qquad (3.6)$$

where $\mathcal{D}$ represents all longitudinal history of the individuals in the sample, information about their survival times, and covariates. Note that (3.6) follows the generic notation (without defining an association structure) of the full joint distribution presented in (3.1).

Our proposal is to employ the inherent sequential nature of the Bayesian methodology to learn from the initial data, summarised in

the joint posterior distribution $\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D}_1)$, and use new information $\mathcal{D}_2$ to update the joint posterior distribution of $(\boldsymbol{b}, \boldsymbol{\theta})$ in order to predict/estimate the relevant survival/longitudinal outcomes.

The use of the Algorithm 1, presented in Chapter 2, for updating the joint posterior distribution of $(\boldsymbol{b}, \boldsymbol{\theta})$ from $\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D}_1)$ to compute $\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$ is not direct, since it does not consider random effects. However, there are extensions of the IBIS algorithm that incorporates random effects, such as generalised linear mixed models (Fan *et al.*, 2008) and longitudinal data with model uncertainty (McGree *et al.*, 2016), but none of them is developed within the context of joint models. For these reasons, we propose in this thesis an extension of the IBIS algorithm accommodated to the random effects in joint models (3.1).

One of the challenges when applying SMC methods, particularly the IBIS algorithm, for static models with random effects is the use of the marginal (or integrated) likelihood function of $\boldsymbol{\theta}$, $f(\mathcal{D} \mid \boldsymbol{\theta})$. As we are interested in applying the standard IBIS algorithm, it will be necessary to obtain the marginal likelihood function of $\boldsymbol{\theta}$ integrating out the random effects of the complete likelihood function (more details below). An advantage of this approach is the reduction of computational effort, since integrating out the random effects is faster than updating them at each iteration (probably with some MCMC algorithm).

For instance, let $y_{i,1:n_i}$ be the vector of size n_i of longitudinal measurements for individual i and T_i their observed event time obtained as the minimum between the true event time T_i^* and the

right censoring time C_i, i.e., $T_i = \min(T_i^*, C_i)$. The event indicator $\delta_i = I\left(T_i^* \leq C_i\right)$ takes the value 1 if the observed time corresponds to a true event time, and 0 otherwise. Here $y_{i,1:n_i}$ is associated to the longitudinal process $\boldsymbol{y}$ and the observed (censored or not) event time (t_{i,n_i}, δ_i) to the time-to-event process $\boldsymbol{s}$. The marginal likelihood function of $\boldsymbol{\theta}$ for the observations $\mathcal{D}_{in_i} = \left[y_{i,1:n_i}, (t_{i,n_i}, \delta_i)\right]^\top$ of the individual i from the beginning of the study until time t_{i,n_i} is obtained by integrating out their random effects $\boldsymbol{b}_i$ as follows:

$$f\left(\mathcal{D}_{in_i} \mid \boldsymbol{\theta}\right) = \int f\left(y_{i,1:n_i}, (t_{i,n_i}, \delta_i) \mid \boldsymbol{b}_i, \boldsymbol{\theta}\right) f\left(\boldsymbol{b}_i \mid \boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{b}_i. \qquad (3.7)$$

Unfortunately, the integral in (3.7) is usually analytically intractable and some numerical integration method is required to approximate it, as pointed out in Gerber and Chopin (2015) and McGree *et al.* (2016).

Among the various options of integration methods (Niederreiter, 2003), we will focus on the two most popular approaches: standard Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods. In general terms, all integration methods based on the Monte Carlo approach approximate (3.7) by:

$$\hat{f}\left(\mathcal{D}_{in_i} \mid \boldsymbol{\theta}\right) = \frac{1}{L} \sum_{l=1}^{L} f\left(y_{i,1:n_i}, (t_{i,n_i}, \delta_i) \mid \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}\right), \qquad (3.8)$$

where $\boldsymbol{b}_i^{(l)}$ is simulated from $f\left(\boldsymbol{b}_i \mid \boldsymbol{\theta}\right)$ for $l = 1, \ldots, L$ (integration nodes). Specifically, for the standard MC and QMC methods employed we simulate values from $f\left(\boldsymbol{b}_i \mid \boldsymbol{\theta}\right)$ based on the well-known inverse cumulative distribution function method (Devroye, 1986).

Firstly, we consider the standard MC integration method for the approximation (3.8). Let $F(\boldsymbol{b}_i \mid \boldsymbol{\theta})$ be the cumulative distribution function of $f(\boldsymbol{b}_i \mid \boldsymbol{\theta})$ with inverse function denoted by $F_{\boldsymbol{b}_i}^{-1}$, then the random variable $\boldsymbol{b}_i = F_{\boldsymbol{b}_i}^{-1}(\boldsymbol{U} \mid \boldsymbol{\theta})$ has cumulative distribution function $F(\boldsymbol{b}_i \mid \boldsymbol{\theta})$, where $\boldsymbol{U}$ is a random variable that follows a standard uniform distribution. In this scheme, we simply have to simulate L integration nodes $(\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(L)})$ with the same dimension of $\boldsymbol{b}_i$ from a standard uniform distribution and then get $\boldsymbol{b}_i^{(l)}$ from the calculation $F_{\boldsymbol{b}_i}^{-1}(\boldsymbol{u}^{(l)} \mid \boldsymbol{\theta})$ for $l = 1, \ldots, L$. Finally, we calculate $\hat{f}(\mathcal{D}_{in_i} \mid \boldsymbol{\theta})$ as in (3.8) from the L vectors of $\boldsymbol{b}_i$ simulated.

As a second approach, we have the QMC methods, based on quasi-random (also known as low discrepancy) sequences and not on the simulation of integration nodes from a standard uniform distribution. Examples of such sequences include Halton (Halton, 1960), Sobol' (Sobol', 1967), and Faure (Faure, 1982). The main advantage of this strategy is the high uniformity in distribution, since these quasi-random sequences are more uniform than random sequences because randomly chosen points tend to clump (Morokoff and Caflisch, 1994). In the case of Monte Carlo integration, this feature leads to faster convergence (in the sense of needing fewer integration nodes) when compared to the standard MC method. The reason for this is that many scenarios require more uniformity than randomness. In this thesis, we only consider the Halton (1960) proposal, which is based on finer and finer prime-based divisions of sub-intervals of unit interval $[0, 1]$. It is important to bear in mind that the selected integration procedure must ideally be as efficient as possible with a low computational cost, since the marginal

likelihood function of $\boldsymbol{\theta}$ in (3.8) will be evaluated many times (in the *weighting* and *moving* steps). After calculating (3.8) whenever required, the standard IBIS algorithm can be used again.

Remembering that the main motivation of this thesis is personalised medicine and we have assumed that the individualised/personalised information of each patient is modelled through random effects $\boldsymbol{b}$. Therefore, we are interested in updating the joint posterior distribution of $(\boldsymbol{b}, \boldsymbol{\theta})$ in (3.6), which can be rewritten as:

$$\pi(\boldsymbol{b}, \boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2) = \pi(\boldsymbol{b} \mid \mathcal{D}_1, \mathcal{D}_2, \boldsymbol{\theta}) \, \pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2). \qquad (3.9)$$

Note that the IBIS algorithm, including the integration part, produces the second posterior distribution in (3.9), $\pi(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2)$. The component $\pi(\boldsymbol{b} \mid \mathcal{D}_1, \mathcal{D}_2, \boldsymbol{\theta})$ is the conditional posterior distribution of the random effects given $\boldsymbol{\theta}$, and its (proportional) expression is obtained by:

$$\pi\big(\boldsymbol{b} \mid \mathcal{D}_1, \mathcal{D}_2, \boldsymbol{\theta}\big) \propto f\big(\mathcal{D}_1, \mathcal{D}_2 \mid \boldsymbol{b}, \boldsymbol{\theta}\big) f\big(\boldsymbol{b} \mid \boldsymbol{\theta}\big). \qquad (3.10)$$

The conditional posterior distribution of $\boldsymbol{b}$ in (3.10) can be updated using a standard Metropolis-Hastings algorithm, as in Rizopoulos *et al.* (2010), based on posterior samples of $\boldsymbol{\theta}$. This additional procedure is an extension of SMC methods for static parameters within the framework of Bayesian joint models as defined in (3.1). We have named this additional procedure in the IBIS approach as a *personalise* step. Algorithm 2 shows our sequential update procedure for Bayesian joint models of longitudinal and time-to-event data.

Algorithm 2: Sequential Monte Carlo algorithm for joint models

1 INITIALISING: draw $\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta} \mid \mathcal{D}_1)$ and set $w^{(k)} \leftarrow 1/K$, $k = 1, \dots, K$.

2 WEIGHTING: from new data $\mathcal{D}_2$, calculate

$$\widetilde{w}^{(k)} \leftarrow \hat{f}\left(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta}^{(k)}\right) w^{(k)},$$

and normalise the weights $w^{(k)} \leftarrow \dfrac{\widetilde{w}^{(k)}}{\sum_{l=1}^{K} \widetilde{w}^{(l)}}, \quad k = 1, \dots, K.$

if $(ESS < K_T)$ **then**

3   RESAMPLING: draw $\left(\widetilde{\boldsymbol{\theta}}^{(1)}, \dots, \widetilde{\boldsymbol{\theta}}^{(M)}\right)$ from $\left(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)}\right)$ with probabilities proportional to the normalised weights $(M \leq K)$. Update $w^{(r)} \leftarrow 1/M, \quad r = 1, \dots, M$.

4   MOVE: draw $\check{\boldsymbol{\theta}}^{(r)}$ from a Metropolis-Hastings kernel of invariant distribution $\pi\left(\boldsymbol{\theta} \mid \mathcal{D}_1, \mathcal{D}_2\right), \quad r = 1, \dots, M$. Update $\boldsymbol{\theta}^{(r)} \leftarrow \check{\boldsymbol{\theta}}^{(r)}, \quad r = 1, \dots, M$ and $K \leftarrow M$.

end

5 PERSONALISE: simulate $\boldsymbol{b}^{(k)} \sim \pi\left(\boldsymbol{b} \mid \mathcal{D}_1, \mathcal{D}_2, \boldsymbol{\theta}^{(k)}\right), \quad k = 1, \dots, K.$

If new data available, return to WEIGHTING step.

The dynamic update mechanism presented in the Algorithm 2 can also be extended to the framework of utility functions in the presence of model uncertainty (Drovandi *et al.*, 2013, 2014; McGree *et al.*, 2016; Ryan *et al.*, 2016).

As introduced in Section 2.2, the main advantage of the sequential Monte Carlo approach in model selection is the approximate update of the normalising constant, leading to an estimate of posterior model probabilities, with negligible additional computational effort. In Section 2.1, we discussed that one of the main difficulties of the Bayesian approach is the computation of the normalising constant $m(\mathcal{D})$. In the case of general sequential Monte Carlo algorithms,

this constant can be approximated as the product of the sum of the unnormalised weights of each iteration (Del Moral *et al.*, 2006). In our sequential strategy the normalising constant can also be updated using this same idea. Our initial analysis using the data set $\mathcal{D}_1$ is non-sequential (MCMC approach) and this is a key difference with respect to the usual sequential methods. Hence, the constant $m(\mathcal{D}_1)$ must be previously calculated. From this assumption and considering a new set of observations $\mathcal{D}_2$, the sequential updating of the "new" normalising constant, which is the basis for updating the Bayes factor, is obtained by:

$$m(\mathcal{D}_1, \mathcal{D}_2) = m(\mathcal{D}_1) \frac{m(\mathcal{D}_1, \mathcal{D}_2)}{m(\mathcal{D}_1)} = m(\mathcal{D}_1)\, m(\mathcal{D}_2 \mid \mathcal{D}_1). \qquad (3.11)$$

From the definition of the normalising constant in (2.1) and approaching the marginal likelihood of $\boldsymbol{\theta}$ by (3.8), we can approximate (3.11) as:

$$\begin{aligned}
m(\mathcal{D}_1, \mathcal{D}_2) &\approx m(\mathcal{D}_1) \int \hat{f}(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid \mathcal{D}_1)\, \mathrm{d}\boldsymbol{\theta} \\
&\approx m(\mathcal{D}_1) \sum_{k=1}^{K} \hat{f}\left(\mathcal{D}_2 \mid \mathcal{D}_1, \boldsymbol{\theta}^{(k)}\right) w^{(k)} \\
&= m(\mathcal{D}_1) \sum_{k=1}^{K} \widetilde{w}^{(k)}, \qquad (3.12)
\end{aligned}$$

where $\widetilde{w}^{(k)}$ are the unnormalised weights, for $k = 1, \ldots, K$, as presented in Algorithm 2.

Analogously, if our interest in a general joint scenario is to calculate the predictive distribution for a set of new observations, d_{new}, given

the current data $\mathcal{D}$, then we can approximate it by means of

$$
\begin{aligned}
f(d_{new} \mid \mathcal{D}) &= \int \hat{f}(d_{new} \mid \mathcal{D}, \boldsymbol{\theta}) \, \pi(\boldsymbol{\theta} \mid \mathcal{D}) \, \mathrm{d}\boldsymbol{\theta} \\
&\approx \sum_{k=1}^{K} \hat{f}(d_{new} \mid \mathcal{D}, \boldsymbol{\theta}^{(k)}) w^{(k)}.
\end{aligned}
\tag{3.13}
$$

In order to apply the sequential methodology described in this chapter for joint models, in the next two chapters we will illustrate its use in a simulated example (longitudinal objective) and in a real application (time-to-event objective).

Applying the sequential methodology in simulated data

As we discussed in Section 1.3, there are basically two possible objectives for using joint modelling: (i) the study of a longitudinal measurement and (ii) the study of the time until an event of interest occurs. Recall that for objective (i) time-to-event models are needed in order to consider informative missing data (see Rubin, 1976), while in (ii) longitudinal models are advisable to introduce endogenous time-varying covariates related with the risk of suffering the event (see Section 1.2).

This chapter illustrates the behaviour of the sequential methodology proposed in Chapter 3 for analysing data from a simulated joint model which interest relies on the longitudinal process, and the time-to-event process is defined for modelling an *informative dropout*[1] (Asar *et al.*, 2015; Elashoff *et al.*, 2016). The chapter also discusses a simulation strategy for generating longitudinal data

with informative dropout as well as briefly presents evidence of the importance of joint modelling for analysing longitudinal data with nonignorable missing data.

All computational time analyses of this chapter were performed on a Windows laptop with an Intel(R) Core(TM) i5-3337U 1.80GHz processor, 2 cores, 4 logical processors, 4GB of RAM and 3MB of cache memory, and the time presented is based on the average of 50 runs. All implementations were made on the `R` environment (version 3.4.0) and, in particular for the MCMC approach, we used the `JAGS` software (version 4.2.0) through the `rjags` package (version 4-6).

4.1 Simulating longitudinal data with informative dropout

Informative (or nonrandom) dropout can be seen as a source of missing data which is often related with a MNAR mechanism in the longitudinal process (see Section 1.1). For instance, in a medical context, the reason why a patient leaves the study prematurely can be associated to the target biomarker in a nonignorable way.

In general, dropout can be assessed in two different ways: according to the time in which it occurs or by means of the indicator of dropout in each observed time (Chan, 2016). We will focus here on the second approach, where for each instant of time t of the follow-up of

---

[1]Also called *monotone missing data*, it occurs when an individual leaves the study at some point and never return, and this information cannot be ignored.

individual i we have a dropout indicator event $\delta_i(t)$ (1 if dropout, 0 otherwise). Hence, whenever we refer to time-to-dropout, we understand it as the time until the dropout indicator changes from 0 to 1, i.e., the time at which the individual leaves the study. As the dropout may occur or not at each time, a Bernoulli model seems sensible to represent this process (Satty and Mwambi, 2013). In this context, if an individual remains in the monitoring system until the end of the study, then we say that their time-to-dropout is censored.

In our simulation strategy, we will assume a shared-parameter joint model framework (see Section 3.1) to connect the longitudinal $(y_i(t))$ and the dropout $(\delta_i(t))$ processes. Thus, in brief, our reference joint model for individual i, for $i = 1, \ldots, N$, is given by:

$$
\begin{aligned}
\big(y_i(t) \mid \mu_i(t), \cdot\big) &\sim \mathcal{P}\big(\mu_i(t), \cdot\big), \\
\big(\delta_i(t) \mid p_i(t)\big) &\sim \mathcal{B}\big(p_i(t)\big), \\
z\big(p_i(t) \mid \mu_i(t), \cdot\big) &= \eta\big(\mu_i(t), \cdot\big).
\end{aligned}
\tag{4.1}
$$

Here $y_i(t)$ denotes the longitudinal variable for individual i at time t which follows a generic probability distribution $\mathcal{P}(\mu_i(t), \cdot)$ with mean $\mu_i(t)$ and other possible parameters. Note that we are not giving any structure for $\mu_i(t)$, but it might be related to some covariates and random effects. In the time-to-dropout process, the random variable $\delta_i(t)$ follows a Bernoulli distribution with parameter $p_i(t)$. The probability of dropout for an individual i at time t, $p_i(t)$, is modelled by some link function $z(\cdot)$ that describes how the mean of the variable $\delta_i(t)$ depends on a linear predictor $\eta(\mu_i(t), \cdot)$ involving $\mu_i(t)$ and other possible covariates and parameters. Some

examples of link functions for the Bernoulli model are logit, probit, and complementary log-log (McCullagh and Nelder, 1989).

The simulation scheme starts with the specification of the number of individuals (N), the number of longitudinal observations (n_i) for each individual i of the sample and their respective measurement times $(t_{i,1:n_i})$, for $i = 1, \ldots, N$. A structure for $\mu_i(t)$ and a for $p_i(t)$ should also be selected, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$, as well as all parameters and hyperparameters of the joint model.

From this preliminary information for each individual i, we start with the simulation of their n_i longitudinal observations from the probability distribution function $\mathcal{P}$ with mean $\mu_{i,j} = \mu_i(t_{i,j})$, for $j = 1, \ldots, n_i$.

In the next step, we generate dropout indicators for each individual. We sequentially calculate the probability that an individual i leaves the study at time $t_{i,j}$ by means of $p_{i,j} = p_i(t_{i,j})$, for $j = 1, \ldots, n_i$. In each iteration of this process, we compare the value of $p_{i,j}$ with a drawn value u from a standard uniform distribution, which is the strategy adopted by the inverse cumulative distribution function method (see Section 3.2). Hence, if $u < p_{i,j}$ then the dropout is not produced at time $t_{i,j}$ (i.e., $\delta_{i,j} = \delta_i(t_{i,j}) = 0$) and we continue with this comparison at next time $t_{i,j+1}$. Otherwise, $\delta_{i,j} = 1$ and consequently the last longitudinal observation of the individual i occurs at time $t_{i,j}$.

The key idea behind this strategy is that the mean $\mu_i(t)$ for an individual i at time t increases or decreases the probability that this

individual continues in the study.

Algorithm 3 displays the simulation procedure described above.

Algorithm 3: Scheme for simulating longitudinal data with informative dropout

---

1 SPECIFICATION:

N # number of individuals.

n_i # number of observations for individual i, for $i = 1, \ldots, N$.

$t_{i,1:n_i}$ # n_i measurement times for individual i, for $i = 1, \ldots, N$.

Values for parameters and hyperparameters of the joint model.

2 SIMULATE ALL OBSERVATIONS OF ALL INDIVIDUALS:

$$y_i(t_{i,j}) \sim \mathcal{P}\big(\mu_{i,j}, \cdot\big), \quad \text{for } j = 1, \ldots, n_i, \ i = 1, \ldots, N,$$

where $\mathcal{P}$ is a generic probability distribution with mean $\mu_{i,j}$ at time $t_{i,j}$.

3 GENERATE THE DROPOUT TIME OF EACH INDIVIDUAL:

for $(i : 1 \text{ to } N)$ **do**

    $\delta_i(t_{i,1:n_i}) \leftarrow 0$.

    for $(j : 1 \text{ to } n_i)$ **do**

        Calculate $p_{i,j}$ and draw $u \sim \mathcal{U}(0, 1)$.

        if $(u > p_{i,j})$ **then**

            $\delta_{i,j} \leftarrow 1$.

            $n_i \leftarrow j$.

            Break.

        end

    end

end

4 RETURN: $y_{i,j}, t_{i,j}, \delta_{i,j}, j = 1, \ldots, n_i, i = 1, \ldots, N$.

Although the procedure presented in Algorithm 3 simulates dropout data, its extension to simulate *intermittently missing data*[2] is very simple. We just have to remove the commands $n_i \leftarrow j$ and Break in Step 3, thus allowing missing data at specific follow-up times.

4.2   Simulated data

To illustrate the simulation procedure described in the previous section, we will consider a joint model with a longitudinal linear mixed-effects model (Verbeke and Molenberghs, 2000; Galecki and Burzykowski, 2013) and a dropout indicator event modelled as in (4.1) employing a logit link function for $p_i(t)$. So our joint modelling for individual i, for $i = 1, \ldots, N$, is given by:

$$
\begin{aligned}
\big(y_i(t) \mid \mu_i(t), \sigma\big) &\sim \mathcal{N}\big(\mu_i(t), \sigma^2\big), \\
\big(\mu_i(t) \mid \boldsymbol{b}_i, \boldsymbol{\beta}\big) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, t, \\
\big(\boldsymbol{b}_i \mid \sigma_0, \sigma_1\big) &\sim \mathcal{N}\Big((0,0)^\top, \operatorname{diag}(\sigma_0^2, \sigma_1^2)\Big), \\
\big(\delta_i(t) \mid p_i(t)\big) &\sim \mathcal{B}(p_i(t)), \\
\operatorname{logit}\big(p_i(t) \mid \mu_i(t), \gamma\big) &= \gamma + \alpha\mu_i(t),
\end{aligned}
\qquad (4.2)
$$

where the longitudinal variable $y_i(t)$ at time t is normally distributed with mean $\mu_i(t)$ and variance σ^2, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ are fixed effects coefficients for the intercept and the slope of the mean longitudinal trajectory, respectively, and $\boldsymbol{b}_i = (b_{0i}, b_{1i})^\top$ are the subsequent individual random effects. We assume that these random effects, b_{0i} and b_{1i}, are independent and normally distributed with mean 0 and variances σ_0^2 and σ_1^2, respectively. In the logit link function for $p_i(t)$ we have an intercept coefficient γ and the association between the longitudinal outcome and the dropout event is expressed by the product of α and $\mu_i(t)$ (i.e., the logit of the probability for dropout and the longitudinal mean are linearly associated).

---

[2]Also called *non-monotone missing data*, it occurs when an individual misses particular visits during the course of the longitudinal study and return at later scheduled visits (Tseng *et al.*, 2016).

The simulation setting is given by $N = 100$ individuals and longitudinal times $0, 1, \ldots, 10$. The parameters and hyperparameters of the model are selected as follows: $\beta_0 = 0.3$, $\beta_1 = 0.5$, $\sigma = 2$, $\sigma_0 = 0.3$, $\sigma_1 = 0.3$, $\gamma = -3$, and $\alpha = 0.7$.

Figure 4.1-(a) presents the longitudinal measures without dropouts with the setting described above. Figure 4.1-(b) shows the same simulated data as in Figure 4.1-(a) with some observations removed as a consequence of the incorporation of the dropout process presented in the previous section.

(a) Without dropout.   (b) With informative dropout.
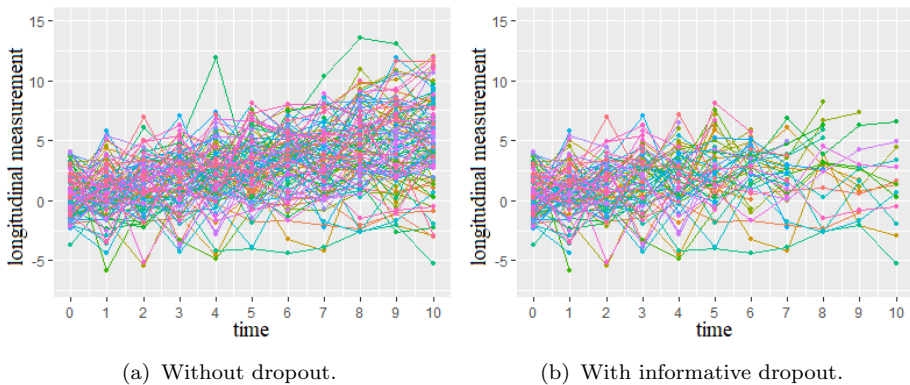
FIGURE 4.1: Longitudinal measurements generated from Algorithm 3 for 100 individuals.

We observe in Figure 4.1-(b) that most individuals leave the study before it ends, i.e., there are few censored data. In addition, there is a (nonrandom) factor that induces individuals to leave the study as their longitudinal measurements increase. In a medical context, this situation could be interpreted when a physician observes high values of a biomarker for a patient (supposing that high values are a sign of disease worsening) and, as a consequence, this patient is

more likely to be taken to another hospital unit (which would be
the reason for leaving the study). A direct comparison between
the two graphics in Figure 4.1 also supports the intuitive idea that
dropout modelling must not be ignored, since at least the common
slope coefficient (β_1) will be poorly estimated. Figure 4.2 shows the
frequency of dropout and censoring times from simulated data in
Figure 4.1-(b).
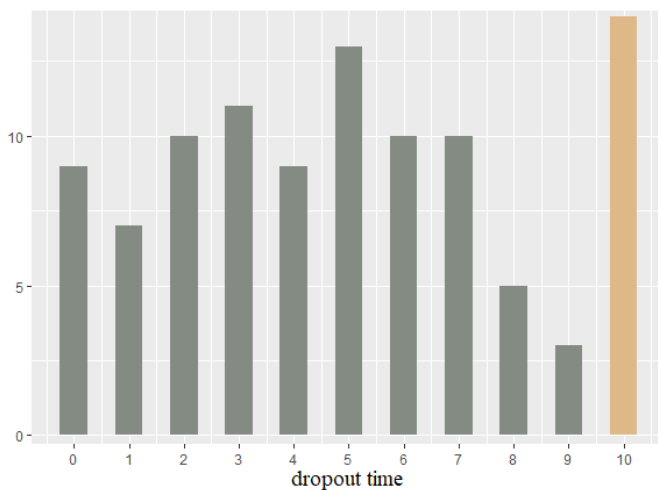
FIGURE 4.2: Frequency of dropout times (gray) and censored
times (burlywood) from simulated data in Figure 4.1-(b).

In the next section, we will use the data generated here to show
the benefits of a Bayesian joint modelling in contrast to a purely
longitudinal analysis when informational dropout happens.

4.3   The benefits of a joint modelling

The joint modelling used in this section is the same that the one
implemented in the simulation mechanism, a shared-parameter joint

model (4.2). All sources of uncertainty, $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma, \sigma_0, \sigma_1, \gamma, \alpha)^\top$, are modelled through vague proper marginal prior distributions with the aim of giving all inferential prominence to the data. In particular, the joint prior distribution of $\boldsymbol{\theta}$ is given by:

$$\pi(\boldsymbol{\theta}) = \pi(\beta_0)\,\pi(\beta_1)\,\pi(\sigma)\,\pi(\sigma_0)\,\pi(\sigma_1)\,\pi(\gamma)\,\pi(\alpha), \qquad (4.3)$$

in which we assume prior independence as a default specification. More specifically, we elicit noninformative marginal prior distributions that provide a wide parametric space for starting the Bayesian process. They are given by:

$$
\begin{aligned}
\pi(\beta_0) = \pi(\beta_1) &= \mathcal{N}(0, 100), \\
\pi(\sigma) &= \mathcal{U}(0, 10), \\
\pi(\sigma_0) = \pi(\sigma_1) &= \mathcal{U}(0, 10), \\
\pi(\gamma) &= \mathcal{N}(0, 100), \\
\pi(\alpha) &= \mathcal{N}(0, 100).
\end{aligned}
\qquad (4.4)
$$

As discussed in Section 2.2, our starting point for assessing the probabilistic behaviour of all relevant quantities derived from $\boldsymbol{\theta}$ is an approximate random sample of size K from its posterior distribution, $\pi(\boldsymbol{\theta} \mid \mathcal{D})$. It may have been computed by using MCMC methods, such as those implemented in `WinBUGS` (Lunn *et al.*, 2000), `JAGS` (Plummer, 2003), or any other acceptable software/package. In this study we previously compared the processing time between `WinBUGS` and `JAGS` software, and opted to do all the analyses using `JAGS`[3], since its execution time was around 40% quicker than `WinBUGS`.

---

[3]The code in `JAGS` of the joint model (4.2) with the marginal prior distributions of $\boldsymbol{\theta}$ specified above is available in Appendix A.

After some preliminary testing, the minimum MCMC configuration to achieve convergence based on the potential scale reduction factor (Gelman and Rubin, 1992; Brooks and Gelman, 1998) and the effective number of independent simulation draws (Kass *et al.*, 1998) is given by three Markov chains with 200000 iterations after a burn-in period of 50000 iterations. The effective iterations are thinned by storing every 400th iteration in order to decrease autocorrelation in the sample. From this MCMC configuration, $K = 3 \times 200000/400 = 1500$. Table 4.1 summarises the marginal posterior distribution of the parameters and hyperparameters of the model.

$\boldsymbol{\theta}$	True	Mean	SD	2.5%	50%	97.5%
β_0	0.3	0.313	0.142	0.039	0.316	0.585
β_1	0.5	0.454	0.058	0.348	0.453	0.578
σ	2.0	2.005	0.063	1.885	2.003	2.129
σ_0	0.3	0.315	0.191	0.014	0.304	0.727
σ_1	0.3	0.321	0.044	0.243	0.321	0.418
γ	-3.0	-2.924	0.277	-3.498	-2.916	-2.416
α	0.7	0.646	0.120	0.433	0.640	0.894

TABLE 4.1: Posterior summaries of the parameters and hyperparameters of the joint model (4.2) using `JAGS`.

From Table 4.1 it is easy to see that the mean value and the 95% credible interval of each parameter are in agreement with the true value of the parameter. The biggest difference between the true and the estimated values is in the common slope coefficient β_1, but within the range of the expected. Simulation in joint models is a very unknown subject that undoubtedly could help to better understand the probabilistic behaviour of them.

In order to illustrate the relevance of the use of joint models in cases of informative dropout, we also propose a purely longitudinal model for this same set of data. To maintain consistency with the joint analysis developed before, the longitudinal model does not contain any element for modelling dropout. Consequently, it is defined as follows:

$$
\begin{aligned}
\big(y_i(t) \mid \mu_i(t), \sigma\big) &\sim \mathcal{N}\big(\mu_i(t), \sigma^2\big), \\
\big(\mu_i(t) \mid \boldsymbol{b}_i, \boldsymbol{\beta}\big) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, t, \\
\big(\boldsymbol{b}_i \mid \sigma_0, \sigma_1\big) &\sim \mathcal{N}\left(\big(0,0\big)^\top, \operatorname{diag}\big(\sigma_0^2, \sigma_1^2\big)\right).
\end{aligned}
\tag{4.5}
$$

We also assume prior independence and select the same marginal prior distributions as in (4.4) for the common parameters and hyperparameters that the model (4.5) shares with (4.4). From the MCMC configuration previously presented, Table 4.2 displays a descriptive summary of the approximate posterior distribution for the parameters and hyperparameters of this longitudinal model.

$\boldsymbol{\theta}$	True	Mean	SD	2.5%	50%	97.5%
β_0	0.3	0.424	0.142	0.154	0.426	0.706
β_1	0.5	0.357	0.057	0.242	0.356	0.476
σ	2.0	2.008	0.063	1.892	2.006	2.143
σ_0	0.3	0.303	0.194	0.011	0.287	0.709
σ_1	0.3	0.303	0.042	0.224	0.301	0.391

TABLE 4.2: Posterior summaries of the parameters and hyperparameters of the longitudinal model (4.5) using `JAGS`.

Based on the true value of each parameter and hyperparameter, the longitudinal analysis that ignores the missing data leads to different results for the common intercept and slope coefficients, β_0 and β_1, but maintain similar values for the three standard deviations, σ,

σ_0, and σ_1. In particular, the expected common elements of the line derived from the longitudinal analysis are $0.424 + 0.357t$ and this regression line clearly produces biased results, while in the case of the joint modelling the resulting regression line is $0.313 + 0.454t$, which leads to a much more adequate approximation to the "complete" longitudinal mean behaviour as shown in Figure 4.1-(a).

So far, we have done a non-sequential joint analysis of longitudinal and time-to-event data with all the available observations up to this now, where the running time using `JAGS` software was about 53 minutes.

We will illustrate in the next section a situation in which new observations become available and it is necessary to sequentially update the inferential process.

4.4   Sequential inference

In this section, we will exemplify the performance of our sequential Monte Carlo proposal for dealing with joint models using the generated data set ($\mathcal{D}$) in Section 4.2 and the modelling presented in the previous section.

Imagine that new data are available in two different stages. In first place, we have a new individual ($i = 101$) who has finished the study with eleven longitudinal measurements. Next, an individual ($i = 38$) originally in the study (with only one observation) has four new longitudinal measurements. Obviously, at each stage we also know

that both individuals have not experienced the *dropout* event before the last observed longitudinal measure. Figure 4.3 shows the update of Figure 4.1-(b) with the new data highlighted. The main reason for using sequential methods in these stages of inferential updating is to avoid recalculation of MCMC methods that in this study would probably take another 53 minutes through `JAGS` software.
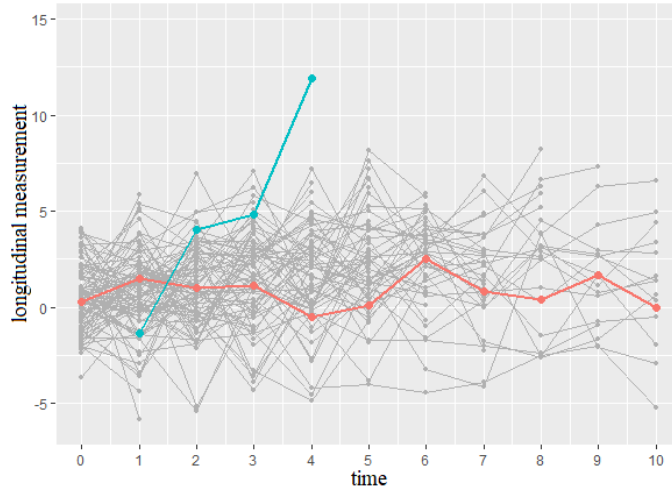
FIGURE 4.3: Observations of the individuals from the initial study (gray), eleven observations for a new individual (guava), and four new observations for an individual originally in the study (blue).

Firstly, we consider that only the new individual is added into the study. In first step (*initialise*) of Algorithm 2, we draw a posterior sample of size $K = 1500$ for each component of $\boldsymbol{\theta}$ obtained from the initial data, $\pi(\boldsymbol{\theta} \mid \mathcal{D})$, and initialise all weights with value 1.

In second step (*weighting*), we compute the incremental importance weights (2.5) from the marginal likelihood function of $\boldsymbol{\theta}$, $f\left(y_{i,1:n_i}, \delta_{i,1:n_i} \mid \boldsymbol{\theta}\right)$, obtained after integrating out the random effects (see Section 3.2). The integral, generally written as in (3.7),

has usually no analytical solution and then some integration method is required. Hence, for each particle k of each component of $\boldsymbol{\theta}$, for $k = 1, \ldots, K$, we approximate $f\left(y_{i,1:n_i}, \delta_{i,1:n_i} \mid \boldsymbol{\theta}^{(k)}\right)$ by means of $\hat{f}\left(y_{i,1:n_i}, \delta_{i,1:n_i} \mid \boldsymbol{\theta}^{(k)}\right)$ calculated by:

$$\frac{1}{L} \sum_{l=1}^{L} f\left(y_{i,1:n_i} \mid \mu_i^{(l,k)}(t_{1:n_i}), \sigma^{(k)}\right) f\left(\delta_{i,1:n_i} \mid p_i^{(l,k)}(t_{1:n_i})\right), \quad (4.6)$$

with $\mu_i^{(l,k)}(t_{1:n_i})$ and $p_i^{(l,k)}(t_{1:n_i})$ described by:

$$\mu_i^{(l,k)}(t_{1:n_i}) = \beta_0^{(k)} + b_{0i}^{(l)} + \left(\beta_1^{(k)} + b_{1i}^{(l)}\right) t_{i,1:n_i}, \quad (4.7)$$

$$p_i^{(l,k)}(t_{1:n_i}) = \frac{\exp\left[\gamma + \alpha \mu_i^{(l,k)}(t_{1:n_i})\right]}{1 + \exp\left[\gamma + \alpha \mu_i^{(l,k)}(t_{1:n_i})\right]}. \quad (4.8)$$

Each component of $\boldsymbol{b}_i^{(l)} = \left(b_{0i}^{(l)}, b_{1i}^{(l)}\right)^{\top}$ is simulated from $\mathcal{N}\left(0, \sigma_0^{2(k)}\right)$ or $\mathcal{N}\left(0, \sigma_1^{2(k)}\right)$, respectively, for $l = 1, \ldots, L$. From joint model (4.2), the longitudinal contribution of the new individual i in (4.6) is the product of normal densities evaluated at observations $y_{1:n_i}$, expressed by:

$$f\left(y_{1:n_i} \mid \mu_i(t_{1:n_i}), \sigma\right) = \prod_{j=1}^{n_i} \left\{ \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_{ij} - \mu_i(t_{i,j})\right)^2\right] \right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n_i}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\|y_{i,1:n_i} - \mu_i(t_{i,1:n_i})\right\|^2\right], \quad (4.9)$$

where $\|v_{1:n_i}\|^2 = \sum_{j=1}^{n_i} v_j^2$ represents the Euclidean distance.

On the other hand, the time-to-dropout contribution of this same individual i in (4.6) is described by the product of Bernoulli densities

evaluated at observations $\delta_{i,1:n_i}$, given by:

$$f\left(\delta_{i,1:n_i} \mid p_i(t_{i,1:n_i})\right) = \prod_{j=1}^{n_i} p_i(t_{i,j})^{\delta_{i,j}} \left[1 - p_i(t_{i,j})\right]^{1-\delta_{i,j}}$$

$$= p_i(t_{i,n_i})^{\delta_{i,n_i}} \prod_{j=1}^{n_i} \left[1 - p_i(t_{i,j})\right]^{1-\delta_{i,j}}. \quad (4.10)$$

The simulation strategy of $\boldsymbol{b}_i$ is based on the Monte Carlo approach, and depending on the computational method employed the processing time for a good approximation of (4.6) may differ considerably. As commented in Section 3.2, we will compare the performance of Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods through a study based on different numbers of integration nodes to approximate the marginal likelihood function of $\boldsymbol{\theta}$ in (3.7). Since $f\left(y_{i,1:n_i}, \delta_{i,1:n_i} \mid \boldsymbol{\theta}\right)$ is directly related to the normalised weights (see (2.5)), we use the distribution of these weights to evaluate the integration methods.

We will approximate the distribution of the normalised weights with a large number of integration nodes ($L = 50000$) and this distribution will be our reference approximation. The approximations with both methods are similar. Figure 4.4 shows the performance of both integration methods for the calculation of the distribution of the normalised weights as well as the reference distribution from the 11 observations of the individual 101.

Clearly, QMC is the method that needs less nodes (about 100) to get a good approximation of the distribution of normalised weights. MC method requires around 750 nodes to reach an approximation
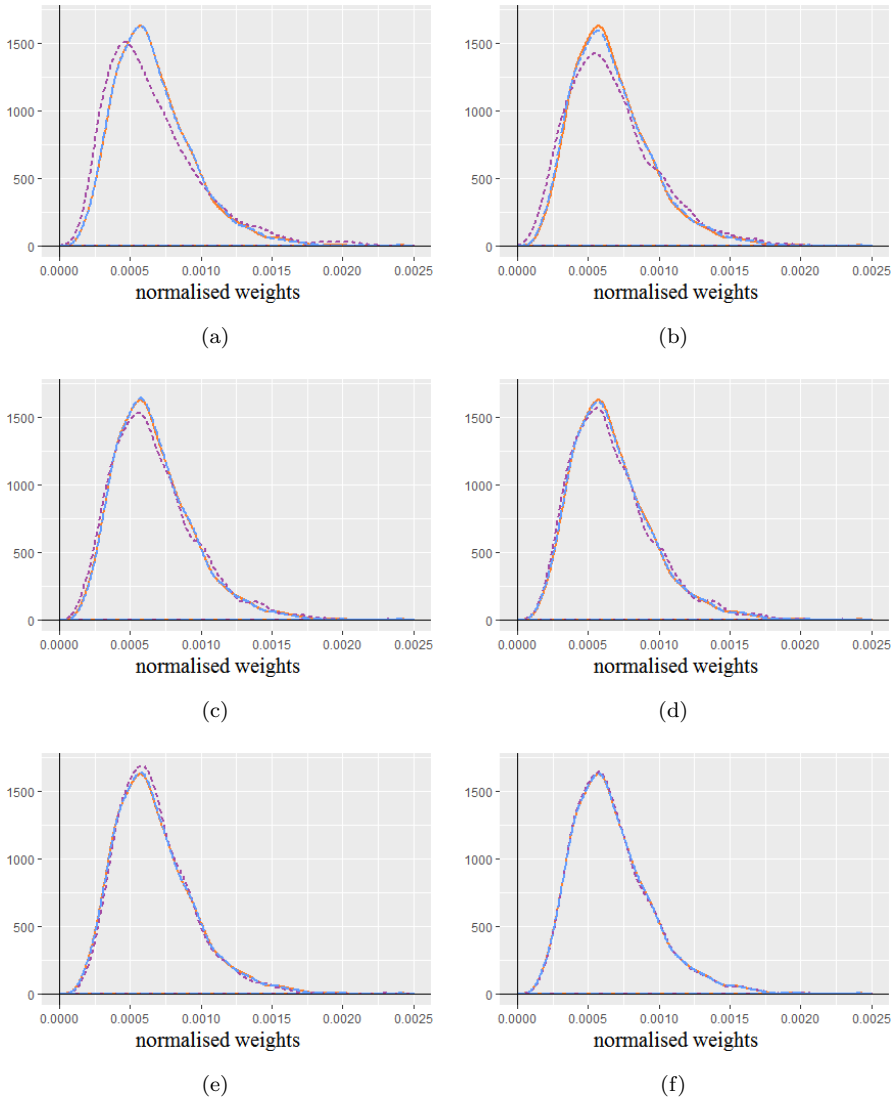
FIGURE 4.4: Distribution of the approximate normalised weights obtained by Monte Carlo (purple dashed line) and quasi-Monte Carlo (blue long-dashed line) integration methods, with $L = 50, 75, 100, 250, 500, 750$ (plots (a) to (f) respectively) and a reference distribution (orange solid line) using $L = 50000$.

equivalent to the QMC with 100 nodes. Based on this results, in this simulated study we will approximate the marginal likelihood function of $\boldsymbol{\theta}$ by quasi-Monte Carlo integration using 100 integration nodes.

Still in the *weighting* step, from the (unnormalised) weights calculated by the product of (4.6) and the initial weights, we should normalise them and check the effective sample size (ESS), which is calculated as in (2.6). Taking the typical threshold $K_T = K/2 = 750$ as reference, we have that $ESS = 1279$ is greater than K_T, so we do not trigger the *resampling* and *moving* steps. In practice, this means that these eleven new observations does not significantly alter the first marginal posterior distribution of the parameters and hyperparameters.

Finally, we update the random effects through the *personalise* step. The conditional posterior distribution of the random effects for all individuals, $\pi(\boldsymbol{b} \mid \mathcal{D}, y_{i,1:n_i}, \delta_{i,1:n_i}, \boldsymbol{\theta})$, is written as in (3.10). However, recalculation of the random effects of individuals originally belonging to the study is unnecessary, since $\boldsymbol{\theta}$ did not need to be updated after inclusion of the new individual i. Therefore, this update reduces to the calculation of the conditional posterior distribution of the random effects for the individual i, expressed by $\pi(\boldsymbol{b}_i \mid y_{i,1:n_i}, \delta_{i,1:n_i}, \boldsymbol{\theta})$ and calculated by means of a standard Metropolis-Hastings algorithm, as indicated in Section 3.2.

The computational time for this sequential procedure was 30 seconds. However, it is important to highlight that we have not entered

in the *resampling* and *moving* steps, and so the processing is much faster.

Imagine that now we want to incorporate the four new longitudinal observations of the individual 38 to the inferential process. In this new update, we proceed from the *weighting* step using the previous samples as the initial set of particles. We compute the incremental importance weights in (2.5) by multiplying the normalised weights $w^{(k)}$, for $k = 1, \ldots, 1500$, stored in the previous update by the marginal approximation

$$\frac{1}{L} \sum_{l=1}^{L} \left[\frac{f\left(y_{i,1:n_i+g} \mid \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}^{(k)}\right) f\left(\delta_{i,1:n_i+g} \mid \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}^{(k)}\right)}{f\left(y_{i,1:n_i} \mid \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}^{(k)}\right) f\left(\delta_{i,1:n_i} \mid \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}^{(k)}\right)} \right], \quad (4.11)$$

where $g = 4$ is the number of new longitudinal observations for individual i. The longitudinal and time-to-dropout components are the same as (4.9) and (4.10), respectively, and the random effects $\boldsymbol{b}_i^{(l)}$ are simulated as in (4.6). Hence, we obtain the updated (unnormalised) weights from the product of (4.11) and the normalised weights after adding all the information of the individual 101 (first update), for $k = 1, \ldots, 1500$. We also update the normalised weights and calculate the ESS, which results in 742. Note that now the effective sample size is less than $K_T = 750$, so the update procedure requires the *resampling* and *moving* steps.

The *resampling* step is performed using the multinomial selection scheme presented in Section 2.2, where the number M of "new" particles for each parameter and hyperparameter is equal to 500 (one-third of the total number of initial particles)[4].

After the particles have been resampled, we trigger the *moving* step. At this stage we move once each resampled particle according to the acceptance probability of the Metropolis-Hastings algorithm. The independent proposal distribution for each new set of particles (composed of a particle of each component of $\boldsymbol{\theta}$) is a multivariate normal distribution with mean vector and diagonal of the variance-covariance matrix (off-diagonal is equal to zero, since the particles are independent of each other) given by:

$$\hat{\mathbf{E}} = \sum_{k=1}^{K} w^{(k)} \boldsymbol{\theta}^{(k)} \qquad \text{and} \qquad \hat{\mathbf{V}} = \sum_{k=1}^{K} w^{(k)} \left(\boldsymbol{\theta}^{(k)} - \hat{\mathbf{E}} \right)^2, \qquad (4.12)$$

as suggested by Chopin (2002). Note that some parameters and hyperparameters (the standard deviations σ, σ_0, and σ_1) only take positive values, hence assuming normality for the proposal distributions of these components is inappropriate. To circumvent this problem, we apply the logarithmic transformation to these components and so we use a normal setting as in (4.12), which is the default approach for the independent proposal distribution.

Now we must update the random effects $\boldsymbol{b}$ based on the "new" $\boldsymbol{\theta}$ (*personalise* step). Again, we use a Metropolis-Hastings approach to approximate the conditional posterior distribution of $\boldsymbol{b}$. However, this time the random effects of all individuals must be updated, since marginal posterior distributions of $\boldsymbol{\theta}$ were also updated.

In this second update, all the steps of the Algorithm 2 were required and the average processing time was about 5 minutes. Note

---

[4]This number was set from some initial calibration tests.

that in both situations, the reduction of computational time is quite relevant, since JAGS software takes around 53 minutes.

Since all new observations have been incorporated into the study, we have a new inference for the parameters and hyperparameters $\boldsymbol{\theta}$. Table 4.3 displays the marginal posterior expectation and standard deviation of $\boldsymbol{\theta}$ after using the sequential approximation. In order to validate and compare our results, we have also included in this table the posterior summary of $\boldsymbol{\theta}$ (re)using JAGS software and the one obtained in Table 4.1.

$\boldsymbol{\theta}$	True	**Initial** (Table 4.1)		**SMC**		**JAGS**	
		Mean	SD	Mean	SD	Mean	SD
β_0	0.3	0.313	0.142	0.295	0.136	0.296	0.140
β_1	0.5	0.454	0.058	0.474	0.060	0.474	0.060
σ	2.0	2.005	0.063	2.006	0.062	2.006	0.063
σ_0	0.3	0.315	0.191	0.300	0.180	0.301	0.180
σ_1	0.3	0.321	0.044	0.346	0.044	0.347	0.044
γ	-3.0	-2.924	0.277	-3.035	0.284	-3.033	0.285
α	0.7	0.646	0.120	0.678	0.122	0.675	0.122

TABLE 4.3: Marginal posterior expectation and standard deviation of $\boldsymbol{\theta}$ before (**Initial**) and after incorporating all new observations. **SMC**: sequential update from the Algorithm 2. **JAGS**: non-sequential update from JAGS software.

Obviously, all these parameters and hyperparameters explain population aspects and so a few new observations (compared to the amount of initial data) would not drastically alter their estimated values. The estimates using our sequential method and the JAGS software are equivalent, since the differences between the columns **SMC** and **JAGS** are minimal (only in the third decimal place).

We analysed other simulation scenarios based on different parameter and hyperparameter settings. The comparative results between MCMC and SMC approaches are presented in Appendix C. In all cases, the sequential methodology presented a significant reduction of the computational time and maintained the accurate estimation of the marginal posterior distribution of each parameter and hyperparameter of the joint model.

Application in ICU discharge data

We continue in this chapter with the illustration of the use of sequential methods for Bayesian joint models proposed in Chapter 3. We will consider a real study (Rué *et al.*, 2017) focused on patients receiving mechanical ventilation (MV) in intensive care units (ICU).

The goal of Rué *et al.* (2017) was to study the probability of occurring the events *alive discharge* or *death* for mechanically ventilated patients in the ICU by means of two severity biomarkers, sequential organ failure assessment (SOFA) score and an asynchronies index (AI), both measured daily. The SOFA score measures the degree of organ dysfunction through scores of 0 (normal) to 4 (most abnormal) of each of the following six organ systems: respiratory, cardiovascular, renal, coagulation, hepatic, and neurological systems. Patient-ventilator asynchrony appears when the cycle of the ventilator is not simultaneous with the patient's respiratory one. Hence, the AI

is defined as the proportion of asynchronous events among the total number of ventilator cycles. The authors proposed a Bayesian joint model with bivariate longitudinal and competing risks submodels. They observe a great level of uncertainty with regard to the relevance of AI when SOFA information is known. For this reason and in order to better illustrate our proposal, we have decided to consider just the SOFA score as a biomarker related to the events of interest.

Note that there are two competing events of interest in this study (*alive discharge* and *death*), which characterise a special class of survival data, the so-called *competing risks* data (Pintilie, 2006). In our context, a patient can experience one of two possible events of interest. In general, the focus of the study is on the distribution of the time-to-event for one of these events in the presence of the other ones.

The statistical analysis was performed using data[1] from four Spanish ICU (Parc Taulí University Hospital, Sabadell; Hospital Sant Joan

---

de Deu-Fundació Althaia, Manresa; Hospital Central de Asturias, Oviedo, and Complejo Hospitalario Universitario de Granada, Granada) from July 2009 to May 2016.

Next, we will describe this data set with its particular characteristics. Then we will specify an alternative joint model to the one in Rué *et al.* (2017), based on the random-effects approach to joint models and compute the marginal posterior distribution of all its parameters and hyperparameters through `JAGS` software. Finally, we will conclude the chapter with the application of our proposal of sequential updating of the inferential process with new patient information.

All computational time analyses of this chapter were performed on a Windows laptop with an Intel(R) Core(TM) i5-3337U 1.80GHz processor, 2 cores, 4 logical processors, 4GB of RAM and 3MB of cache memory, and the time presented is based on the average of 20 runs. All implementations were made on the `R` environment (version 3.4.0) and, in particular for the MCMC approach, we used the `JAGS` software (version 4.2.0) through the `rjags` package (version 4-6).

5.1   Data description

All patients were followed from the first day in MV until ICU discharge (alive or dead) or day 30 after MV initiation. Of the 139 studied patients, 97 (69.8%) were discharged alive, 28 (20.1%) died, and 14 (10.1%) were administratively censored.

In a descriptive analysis, whenever competing risks data are available we need to use the cumulative incidence curves to describe the probabilities of each event (see Figure 5.1). They are an alternative to the Kaplan-Meier estimates that, in competing risk models, underestimate the subsequent survival probabilities (Kim, 2007; Zhang *et al.*, 2008). Figure 5.1 shows that the first 10 days after MV initiation are critical due to the high rates of *alive discharge* or *death*. In fact, cumulative incidences at day 10 of *alive discharge* and *death* are 42.5% and 14.4%, respectively, while at day 20 are 62.6% and 18.7%.
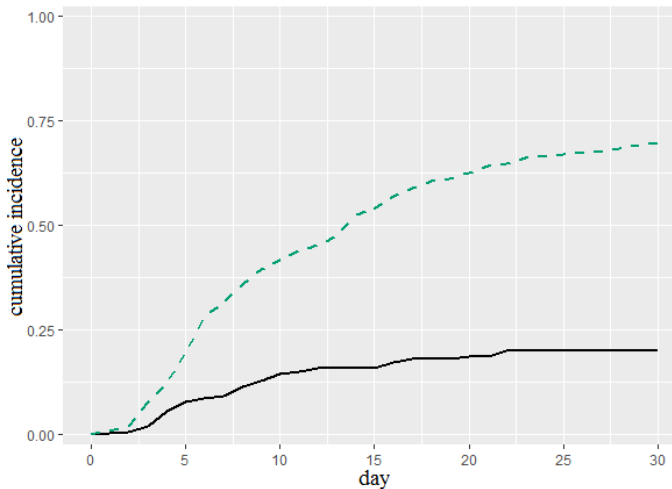
FIGURE 5.1: Cumulative incidence function for *alive discharge* (green dashed line) from the ICU or *death* (black solid line) in the ICU.

With the purpose of later analysing the performance of the sequential methodology for new data, we have eliminated some observations from the database. Therefore, in the initial analysis using JAGS[2] software, we have removed the last three SOFA scores of the

patient 12^3 and all information of the patient 131^4 (who is not con-
sidered in the initial analysis). Hence, the event time for patient
12 will become right-censored in the initial analysis and it will con-
tinue to be right-censored as each "new" SOFA score is observed.
Similarly, the event time for patient 131 will also be right-censored
from the moment that this patient is incorporated into the study
until the inclusion of his/her last SOFA score.

For establishing a sensible longitudinal model for the SOFA
biomarker within the normality, we need to transform it as $SOFA^* = \log(SOFA + 1)$. Adding 1 before applying the logarithm function is
necessary due to the fact that the SOFA score can take values be-
tween 0 and 24 (in our sample the maximum value is 18). Figure 5.2
displays the SOFA (Figure 5.2-(a)) and $SOFA^*$ (Figure 5.2-(b)) tra-
jectories for the patients in the study.

Figure 5.2-(b) shows that $SOFA^*$ trajectories for patients *discharged
alive* are generally lower than those for patients who died. It is
important to remark that in many cases the last observed $SOFA^*$
score of a patient is recorded several days before he/she experienced
one of the events of interest or was administratively censored.

In the next section, we will propose a joint modelling based on a
longitudinal linear mixed submodel for the $SOFA^*$ biomarker and a

---

[2]The code in `JAGS` of the joint model (5.1) and (5.3) with the prior distri-
butions specified above is available in Appendix A.

[3]Patient 12: 71 years old, discharged alive from the ICU at day 6, and its
SOFA scores, from day 1 to 6, were 9, 9, 9, 4, 2, and 2.

[4]Patient 131: 63 years old, died in the ICU at day 5, and its SOFA scores,
from day 1 to 3, were 16, 15, and 15.

(a)   (b)
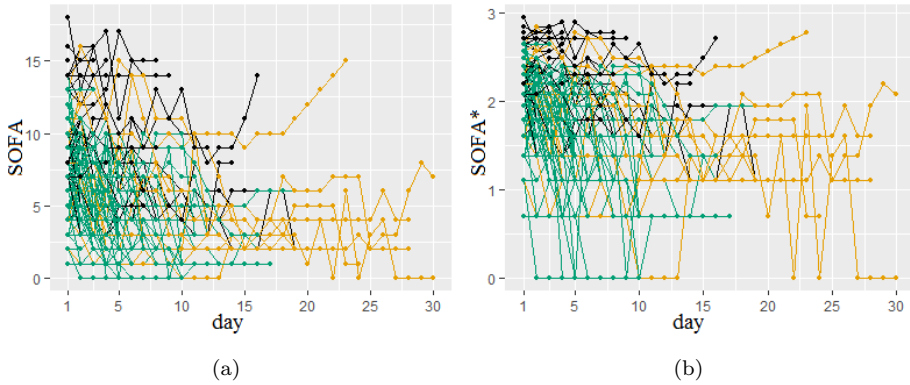
FIGURE 5.2: SOFA (a) and SOFA* (b) longitudinal measurements for patients who discharged alive (green), died (black), and were administratively censored (orange).

survival competing risks submodel for the time to be *alive discharge* and for the time to *death*.

5.2 Modelling and preliminary results

We will assume a fully parametric modelling with the age of the patient as the only baseline covariate in both the longitudinal and the time-to-event submodels. The connection between the longitudinal and competing risks submodels will be through the random-effects approach to joint models described in (3.5). The longitudinal submodel for the ith patient, for $i = 1, \ldots, 139$, is given by:

$$
\begin{aligned}
\big(y_i(t) \mid \mu_i(t), \sigma\big) &\sim \mathcal{N}\big(\mu_i(t), \sigma^2\big), \\
\big(\mu_i(t) \mid \boldsymbol{b}_i, \boldsymbol{\beta}\big) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, t + \beta_2 x_i, \\
\big(\boldsymbol{b}_i \mid \sigma_0, \sigma_1\big) &\sim \mathcal{N}\left((0,0)^{\top}, \operatorname{diag}\big(\sigma_0^2, \sigma_1^2\big)\right),
\end{aligned}
\qquad (5.1)
$$

where $y_i(t)$ expresses the value of SOFA* for the ith patient at time t, which is normally distributed with mean $\mu_i(t)$ and variance σ^2. The parameters β_0 and β_1 are regression coefficients for the intercept and the slope of $\mu_i(t)$, respectively, and b_{0i} and b_{1i} are their subsequent random effects. The random effects b_{0i} and b_{1i} are independent and normally distributed with mean 0 and variances σ_0^2 and σ_1^2, respectively. The covariate x_i represents the age of the ith patient and β_2 is its regression coefficient. We denote by $y_{i,1:n_i}$ the vector of follow-up measurements for the ith patient, where $y_{i,j}$ is the observed SOFA* score at time $t_{i,j}$

The time-to-event modelling is expressed through a cause-specific hazard model (Putter *et al.*, 2007). Let T_{iv}^* the time from the initiation of MV to the occurrence of the event v for the ith patient, where $v = 1$ (*alive discharge* from the ICU) or $v = 2$ (*death* in the ICU); C_i corresponds to the administrative censoring time (day 30 after MV initiation); δ_i is the event indicator, where $\delta_i = 0$ represents censoring for both events, $\delta_i = 1$ indicates that the ith patient is alive at ICU discharge, and $\delta_i = 2$ that the ith patient is dead at ICU; and $T_i = \min\left(T_{i1}^*, T_{i2}^*, C_i\right)$ expresses the observed event time for the ith patient. Hence, for each patient, the cause-specific hazard function measures the instantaneous rate of experiencing a specific event type at time t given that this patient has not experienced any prior event (Kleinbaum and Klein, 2012). So the cause-specific hazard function for event v at time t for the ith patient, for $i = 1, \ldots, 139$, can be written as:

$$h_{iv}(t) = \lim_{\Delta t \to 0} \frac{P\left(t \leq T_i < t + \Delta t, \ \delta_i = v \mid T_i \geq t\right)}{\Delta t}, \quad t \geq 0, \quad (5.2)$$

where $\delta_i = v$ indicates that event v has been selected for the ith patient and Δt is an incremental time. Assuming proportional hazards (Cox, 1972), the cause-specific hazard function (5.2) is rewritten as:

$$h_{iv}(t \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = h_{0v}(t) \exp \left[\gamma_v x_i + \alpha_{0v} b_{0i} + \alpha_{1v} b_{1i} t \right], \quad t \geq 0, \quad (5.3)$$

where $h_{0v}(t) = \lambda_v \nu_v t^{\nu_v - 1}$ represents a Weibull cause-specific baseline hazard function at time t for $v = 1, 2$, with λ_v and ν_v being the scale and shape parameters, respectively. The parameters α_{0v} and α_{1v} quantify the association between the individual characteristics of the SOFA* biomarker and the risk for event v, for $v = 1, 2$. Again, x_i represents the age of the ith patient, where γ_v is its fixed effects coefficient, for $v = 1, 2$.

The vectors of parameters of the joint model (5.1) and (5.3) are defined as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ and $\boldsymbol{\alpha}_v = (\alpha_{0v}, \alpha_{1v})^\top$, and the random effects $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{139})^\top$, where $\boldsymbol{b}_i = (b_{0i}, b_{1i})^\top$ for $i = 1, \ldots, 139$. $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, \sigma_0, \sigma_1, \gamma_1, \gamma_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \nu_1, \nu_2, \lambda_1, \lambda_2)^\top$ is the vector of all parameters and hyperparameters.

We use the random-effects approach to joint models (3.5) and consequently $f(y_{i,1:n_i}, (t_i, \delta_i) \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = f(y_{i,1:n_i} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) f((t_i, \delta_i) \mid \boldsymbol{b}_i, \boldsymbol{\theta})$. The longitudinal contribution $f(y_{i,1:n_i} \mid \boldsymbol{b}_i, \boldsymbol{\theta})$ follows the structure based on normal densities as in (4.9). However, the functional relationship provided by the time-to-event in this conditional joint distribution is different to the one of the simulated study in Chapter 4 due to the time-to-event submodel and the type of random-effects approach employed here. Hence, the competing risks contribution to $f(y_{i,1:n_i}, (t_i, \delta_i) \mid \boldsymbol{b}_i, \boldsymbol{\theta})$ is given by:

$$f\big(t_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}\big) = \prod_{v=1}^{2} h_{iv}\big(t_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}\big)^{I(\delta_i=v)} S_{iv}\big(t_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}\big), \qquad (5.4)$$

where $S_{iv}(t \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \exp\big(-\int_0^t h_{iv}(s \mid \boldsymbol{b}_i, \boldsymbol{\theta})\,\mathrm{d}s\big)$ is the cause-specific survival function for the ith patient at time t, for $v = 1, 2$. By definition, this integral is always one-dimensional, and it can be efficiently approximated using the Q-point Gauss-Legendre quadrature rule, as is done in Armero *et al.* (2016b). We will use this rule with 15 quadrature points in our analysis.

We assume independent and noninformative marginal prior distributions. More specifically, we assign diffuse distributions given by:

$$
\begin{aligned}
\pi(\beta_0) = \pi(\beta_1) = \pi(\beta_2) &= \mathcal{N}(0, 1000), \\
\pi(\sigma) &= \mathcal{U}(0, 100), \\
\pi(\sigma_0) = \pi(\sigma_1) &= \mathcal{U}(0, 50), \\
\pi(\gamma_1) = \pi(\gamma_2) &= \mathcal{N}(0, 1000), \\
\pi(\alpha_{01}) = \pi(\alpha_{02}) &= \mathcal{N}(0, 1000), \\
\pi(\alpha_{11}) = \pi(\alpha_{12}) &= \mathcal{N}(0, 1000), \\
\pi(\nu_1) = \pi(\nu_2) &= \mathcal{G}(0.01, 0.01), \\
\pi(\log(\lambda_1)) = \pi(\log(\lambda_2)) &= \mathcal{N}(0, 1000).
\end{aligned}
\qquad (5.5)
$$

Similarly to the simulated study, here we also did some preliminary tests to define the minimum MCMC configuration to achieve convergence based on the potential scale reduction factor and the effective number of independent simulation draws. In this case, we got three Markov chains with 200000 iterations after a burn-in period of 40000 iterations. The effective iterations were thinned by storing every 400th iteration in order to decrease autocorrelation in the sample.

From this MCMC configuration, $K = 3 \times 200000/400 = 1500$. Table 5.1 displays a descriptive summary of the approximate posterior distribution for the parameters and hyperparameters related to the SOFA* score together with the competing risks process. The last column of this table contains the probability that the corresponding parameter is positive. A probability equal to 0.5 indicates that a positive value of the parameter is equally likely than a negative one.

$\boldsymbol{\theta}$	Mean	SD	2.5%	50%	97.5%	$P(\cdot > 0 \mid \mathcal{D})$
	*Longitudinal process - SOFA**					
β_0	1.844	0.155	1.537	1.844	2.141	1.000
β_1	-0.086	0.009	-0.105	-0.086	-0.067	0.000
β_2	0.005	0.002	0.000	0.005	0.009	0.973
σ	0.311	0.008	0.295	0.310	0.328	—
σ_0	0.407	0.032	0.349	0.405	0.476	—
σ_1	0.067	0.008	0.052	0.066	0.084	—
	Competing risks process - Alive					
γ_1	0.001	0.007	-0.012	0.001	0.016	0.563
α_{01}	-0.203	0.323	-0.846	-0.208	0.468	0.259
α_{11}	-1.012	0.256	-1.582	-0.998	-0.549	0.000
ν_1	1.525	0.151	1.247	1.516	1.834	—
λ_1	0.015	0.009	0.004	0.013	0.040	—
	Competing risks process - Death					
γ_2	0.022	0.017	-0.009	0.021	0.058	0.917
α_{02}	3.367	0.948	1.756	3.307	5.456	1.000
α_{12}	0.745	0.472	-0.173	0.750	1.605	0.943
ν_2	1.172	0.250	0.740	1.157	1.733	—
λ_2	0.002	0.004	0.000	0.001	0.012	—

TABLE 5.1: Posterior summaries of the parameters and hyperparameters of the joint model (5.1) and (5.3) using `JAGS`.

In the longitudinal process, considering the average age (61.82) of the sample patients, we obtain a mean expected value for the first SOFA* of 2.067 (equivalent to 6.905 in the original SOFA scale).

The posterior expectation of the population slope (trend) is slightly decreasing, indicating a reduction of the SOFA* score and possibly an improvement of the patients over time. Age has an associate regression coefficient with a positive posterior expectation, meaning that SOFA* increases with age. The variability (in logarithmic scale) of the measurement error and the intercept random effects have similar magnitudes, while it is smaller for the slope random effects among patients.

In the competing risks process, we identify a clear distinction between the effects of covariates in the risk of *alive discharge* and *death*. For the event *alive discharge*, the age of the patient does not provide relevant information, since its respective parameter γ_1 has a posterior probability of being positive around 0.5. On the other hand, it seems that age of the patient has an incremental positive effect on the risk of *death* in the ICU. Marginal posterior distributions of the association parameters, α_{01} and α_{11}, between the longitudinal and the competing risks process have a predominant negative support indicating that the increase in the SOFA* score implies a reduction of the risk of *alive discharge*. In contrast, the association parameters for the risk of *death* in the ICU, α_{02} and α_{12}, have marginal posterior distributions with mostly positive support and consequently leading to more risk of *death* as the SOFA* score increases. The posterior expectation of the parameters of the Weibull cause-specific baseline hazard function are distinct for each event, but their credibility intervals overlap to a large extent, indicating few differences between them.

In short, the opposite signs of the association parameters between *alive discharge* and *death* is a strong indicative that the SOFA biomarker satisfactorily discriminates the risk of each of these events. In other words, our joint modelling proposal shows that the SOFA biomarker has a direct association with the patients' vital status in the intensive care units considered here.

So far, we have done a non-sequential joint analysis of longitudinal and time-to-event data with all the available observations up to this now, where the running time using `JAGS` software was about 867 minutes.

The next section will be devoted to the application of our sequential proposal presented in Chapter 3, since the observations of patients 12 and 131 (see Section 5.1) will be sequentially introduced in the study and so dynamic estimation will be required.

5.3   Sequential inference

From the point of view of personalised medicine, we are interested in the dynamic estimation of the cumulative incidence function (CIF) either for a specific patient in the study that has provided g new SOFA* scores or for a new patient of the population that could enter the study with g SOFA* scores. It is important to point out that all this longitudinal information also implies survival information with regard to the non-occurrence of none of the events of interest at the time of the last longitudinal observation. In both situations

of data inclusion, we should use a conditional CIF, since the update of this function is conditioned on all current and relevant patient information. In probabilistic terms, the conditional CIF for the ith patient is defined by:

$$P\big(T_i \leq t, \delta_i = v \mid y_{i,1:n_i+g}, T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i, \boldsymbol{\theta}\big), \;\; t > t_{i,n_i+g}, \quad (5.6)$$

where t_{i,n_i+g} is the time at which the last SOFA* score of the ith patient, represented by y_{i,n_i+g}, has been recorded. In particular, the assumption of independence between the longitudinal and the competing risks process, given by the random-effects approach to joint models (3.5), makes the longitudinal contribution irrelevant in the determination of these probabilities. So we can rewrite (5.6) as:

$$P\big(T_i \leq t, \delta_i = v \mid T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i, \boldsymbol{\theta}\big)$$
$$= F_{iv}\big(t \mid T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i, \boldsymbol{\theta}\big), \;\; t > t_{i,n_i+g}. \quad (5.7)$$

From a Bayesian perspective, posterior distribution of the conditional CIF in (5.7) can be derived from the joint posterior distribution of $(\boldsymbol{b}_i, \boldsymbol{\theta})$, $\pi(\boldsymbol{b}_i, \boldsymbol{\theta} \mid \mathcal{D}_i^g, \mathcal{D})$, where $\mathcal{D}_i^g = \big[y_{i,n_i+1:n_i+g}, (t_{i,n_i+g}, \delta_i)\big]^\top$ represents the g new information of the ith patient and $\mathcal{D}$ all the data in the original sample such as longitudinal trajectories, time-to-event times, and covariates, of course all except those that are in $\mathcal{D}_i^g$. In particular, we can obtain the posterior expectation $\mathrm{E}\big(F_{iv}(t \mid T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i, \boldsymbol{\theta}) \mid \mathcal{D}\big)$ by:

$$\int F_{iv}(t \mid T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i, \boldsymbol{\theta}) \, \pi(\boldsymbol{b}_i, \boldsymbol{\theta} \mid \mathcal{D}_i^g, \mathcal{D}) \, \mathrm{d}(\boldsymbol{b}_i, \boldsymbol{\theta}). \quad (5.8)$$

We employ our sequential methodology proposal to update the posterior distribution of $(\boldsymbol{b}_i, \boldsymbol{\theta})$, $\pi\left(\boldsymbol{b}_i, \boldsymbol{\theta} \mid \mathcal{D}_i^g, \mathcal{D}\right)$, and then calculate the posterior expectation in (5.8). Commonly, the integral in (5.8) has no analytical solution and so we should approximate it through a Monte Carlo simulation scheme given by:

$$\frac{1}{L} \sum_{l=1}^{L} F_{iv}\left(t \mid T_i \geq t_{i,n_i+g}, \boldsymbol{b}_i^{(l)}, \boldsymbol{\theta}^{(l)}\right), \tag{5.9}$$

where $\boldsymbol{b}_i^{(l)}$ and $\boldsymbol{\theta}^{(l)}$ are drawn from $\pi\left(\boldsymbol{b}_i, \boldsymbol{\theta} \mid \mathcal{D}_i^g, \mathcal{D}\right)$, for $l = 1, \ldots, L$ and $v = 1, 2$.

Recall that from our sequential proposal we do not directly get the joint posterior distribution $(\boldsymbol{b}_i, \boldsymbol{\theta})$, $\pi\left(\boldsymbol{b}_i, \boldsymbol{\theta} \mid \mathcal{D}_i^g, \mathcal{D}\right)$ (see Sections 3.2). We first have to update the parameters and hyperparameters $\boldsymbol{\theta}$ based on an approximation of the marginal likelihood function of $\boldsymbol{\theta}$ integrating out the random effects $\boldsymbol{b}_i$, and then update these random effects through the *personalise* step. This approximation will be made using the quasi-Monte Carlo method with Halton sequences using 500 integration nodes.

We will consider three sequential situations for the inclusion of the "new" longitudinal observations of the SOFA biomarker for patients 12 and 131. Firstly, we include the fourth SOFA score of the patient 12 and the first one of the patient 131. Secondly, the fifth and second SOFA scores of the patients 12 and 131, respectively, are incorporated into the study. Thirdly, the last SOFA scores of these patients are included in the analysis. Furthermore, at each stage of

data inclusion we also know that both patients did not experience none of the events of interest in this study (*alive discharge* or *death*).

To obtain more accurate in sequential approximations, we will be stricter with the threshold for the effective sample size (ESS) (see (2.6)) in the Algorithm 2. Hence, we will adopt $K_T = 1400$, where $1 \leq K_T \leq K$ (more restrictive) with $K = 1500$ being the size of the marginal posterior samples of $\boldsymbol{\theta}$ (see Section 2.2).

The first sequential update is performed with the data from the first stage. We obtain an $ESS = 1469$, that is greater than K_T. Hence, we do not trigger the *resampling* and *moving* steps. We directly perform the update of the marginal posterior distribution of the random effects (3.10) through the *personalise* step and then calculate the approximation to the posterior conditional CIF by (5.9). This first application of the sequential Monte Carlo method took approximately 13 minutes.

Now we include the second round of the data. Again, we obtain $ESS = 1410$ greater than K_T, so we just update the marginal posterior distribution of the random effects and then calculate the posterior conditional CIF, as before. In this update, we are considering that both patients are already in the study, making the calculation of the importance weight more computationally complicated (see 2.5) than in the case where one of them is a new patient (first update). The average update time was now about 16 minutes.

Finally, by including the third pair of new observations we obtain

$ESS = 1363 < 1400$ that requires the activation of the *resampling* and *moving* steps. Both steps are performed according to the description in Section 2.2. Due to the use of the *resampling* and *moving* steps, the computational time in this update is higher (193 minutes) than the previous two ones, but still much lower than the computational time via MCMC methods from `JAGS` software.

Table 5.2 shows the marginal posterior expectation and standard deviation resulting from this last update. In order to validate and compare our results, we have also included in this table a posterior summary of $\boldsymbol{\theta}$ (re)using `JAGS` software and the one obtained in Table 5.1.

In a general aspect, after the three data inclusions and their respective sequential inferential updating, the interpretation of the parameters and hyperparameters has not changed. The larger difference in the posterior inference with respect to the original data (**Initial**) come from the third data inclusion (**SMC**), where the *resampling* and *moving* steps were required. From Table 5.2, we can also appreciate the equivalence of the results with the **SMC** and **JAGS** approaches (tiny differences from the third decimal place). However, the reduction of the computational cost when using the sequential methodology is indisputably relevant.

As described at the beginning of this section, one of the main interests of personalised medicine in this study is to dynamically estimate the posterior distribution of the conditional CIF given by (5.9). Note that for each sequential update we need to perform the *personalise* step to get the updated marginal posterior distribution

θ	Initial (Table 5.1)		SMC		JAGS	
	Mean	SD	Mean	SD	Mean	SD
*Longitudinal process - SOFA**						
β_0	1.844	0.155	1.851	0.149	1.851	0.154
β_1	-0.086	0.009	-0.088	0.009	-0.088	0.009
β_2	0.005	0.002	0.005	0.002	0.005	0.002
σ	0.311	0.008	0.311	0.007	0.311	0.008
σ_0	0.407	0.032	0.410	0.031	0.410	0.031
σ_1	0.067	0.008	0.069	0.009	0.068	0.009
Competing risks process - Alive						
γ_1	0.001	0.007	0.002	0.007	0.002	0.007
α_{01}	-0.203	0.323	-0.221	0.313	-0.222	0.316
α_{11}	-1.012	0.256	-0.998	0.253	-0.999	0.255
ν_1	1.525	0.151	1.538	0.146	1.537	0.151
λ_1	0.015	0.009	0.015	0.009	0.015	0.009
Competing risks process - Death						
γ_2	0.022	0.017	0.022	0.016	0.022	0.017
α_{02}	3.367	0.948	3.272	0.858	3.278	0.861
α_{12}	0.745	0.472	0.717	0.453	0.716	0.454
ν_2	1.172	0.250	1.176	0.248	1.178	0.249
λ_2	0.002	0.004	0.002	0.003	0.002	0.004

TABLE 5.2: Marginal posterior expectation and standard deviation of θ before (**Initial**) and after incorporating all new observations. **SMC**: sequential update from the Algorithm 2. **JAGS**: non-sequential update from `JAGS` software.

of the random effects and then calculate the posterior conditional CIF. Hence, in order to illustrate the final result of each inferential update, Figure 5.3 presents the individual estimations of the dynamic cumulative incidences for events of interest in each of the sequential updates for patients 12 and 131 in the study.

As previously reported, patient 12 has three SOFA scores (until day 3) in the initial analysis (see Figure 5.3-(a)). These SOFA scores are moderate (all equal to 9), but the next SOFA observations have

(a) Patient 12.

(b) Patient 131.
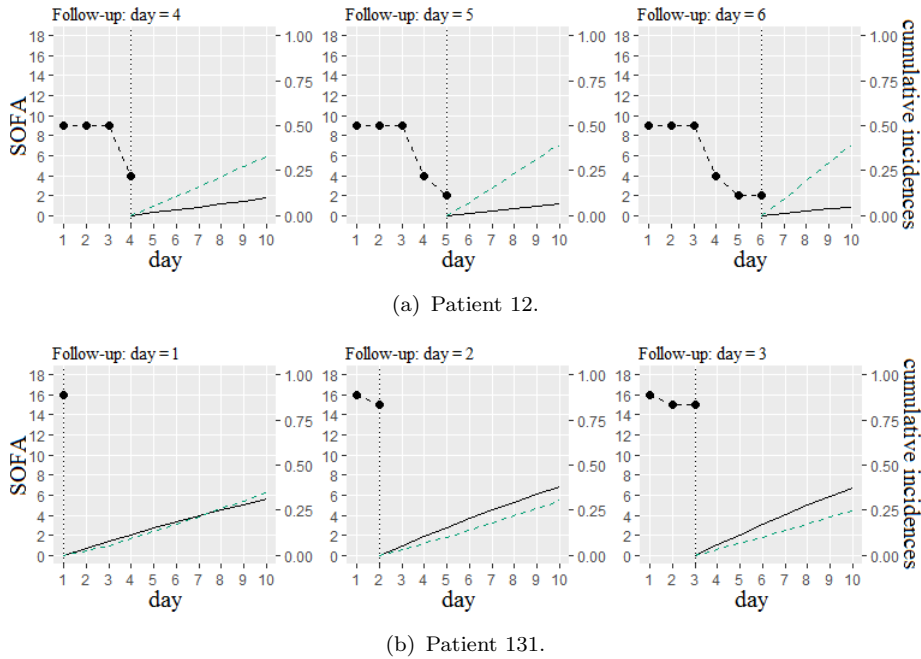
FIGURE 5.3: Individual estimation of the dynamic cumulative incidences for *alive discharge* (green dashed line) from the ICU an *death* (black solid line) in the ICU for patient (a) 12 and (b) 131 in the study. The vertical dotted lines represent the time at which the last SOFA score is observed.

a decreasing trend (4, 2 and 2). The posterior expectation of the conditional CIF of *alive discharge* (green dashed line) increases as new SOFA scores decrease, and in the case of *death* (black solid line) it is low and slowly decreases over time. Patient 12 was discharged alive at day 6. In contrast, patient 131 enters the study with a high SOFA score (16) and it remains high for the two following SOFA observations, both are 15 (see Figure 5.3-(b)). The posterior expectation of the conditional CIF for both events (*alive discharge* and *death*) at day 1 is similar. However, the recurrence of high SOFA scores leads to an increase in the posterior expectation of

the conditional CIF of *death* (black solid line), while the subsequent posterior CIF mean of *alive discharge* progressively decreases.

Final conclusions and future work

In a medical framework, Berzuini *et al.* (1997) emphasise the importance of speeding up the updating of estimates and predictions of interest since the physician usually wants an immediate assessment of the prognostic as new information of the patient becomes available. Hence, our main motivation throughout this thesis has been the optimisation of the inferential update process within the context of new sequential knowledge in a scenario of longitudinal and time-to-event data.

6.1   Conclusions

Inspired in this context, our starting point was to verify the main advantages of a Bayesian joint analysis for longitudinal and time-to-event data. Briefly, we found that the Bayesian joint approach

provides complete inference (longitudinal, time-to-event, and association between both of them), reduces estimation biases, models the informative censoring/dropout, increases statistical efficiency, and conveniently makes predictions of the outcomes (Muthén *et al.*, 2009; Ibrahim *et al.*, 2010; Wang *et al.*, 2012).

In general, Bayesian joint modelling has a high computational cost due to their complexity. This situation gets worse when the relevant information is obtained in a sequential manner and an update of the current knowledge becomes indispensable. In this way, the main contribution of this thesis is to propose and implement dynamic procedures based on sequential Monte Carlo (SMC) methods in joint models. The primary objective is to generate quicker and accurate updated inferences and predictions.

Joint models usually contain different types of random effects, and SMC methods for models with random effects require important adaptations involving analytically intractable integrals. We have circumvented this problem by using integration methods on the Monte Carlo approach. Specifically, we have used and compared the performance of the Monte Carlo and quasi-Monte Carlo integration methods. In addition, we have incorporated to the dynamic procedure the update of the random effects by combining the previously updated parameters and hyperparameters with a standard Metropolis-Hastings algorithm.

We have illustrated the behaviour of our SMC proposal in two different studies. The first one deals with simulated data and has a

longitudinal objective. The second one is based on a real study focused on survival with patients at intensive care units (ICU).

For the first study, we have also proposed a simulation algorithm for generating longitudinal data with informative dropout. This mechanism is essential for linking longitudinal and time-to-dropout processes, since, in general, data generators of this type do not consider a modelling with shared information between both processes. We have generated a data set with this mechanism that has been studied in order to illustrate the performance of our SMC proposal. As an extra bonus, we have also analysed these data from a purely longitudinal point of view, without taking into account the dropout, to emphasise its weaknesses and reinforce the benefits of a joint models approach. Using this generated data set, we have explored in detail the use of our sequential update proposal applied in a joint model with longitudinal objective. In addition, we have compared the posterior distribution of the parameters and hyperparameters obtained from our sequential approach with the one from (non-sequential) MCMC methods via `JAGS` software. The results are essentially the same but the computational time is significantly lower using the SMC approach. Specifically, the reduction was approximately 90% of computational time when it was necessary to update parameters, hyperparameters, and random effects, and about 99% when the update was required only for random effects.

The second study in which we have applied our strategy of sequential updating of the inferential process is a joint model with competing risk events for a real data set. In this study, the main clinical

interest was to connect the information of the sequential organ failure assessment (SOFA) biomarker to the events *alive discharge* and *death* for patients hospitalised in the ICU as well as to dynamically compute for a specific patient the posterior expectation of the conditional cumulative incidence function of each of these events. In this analysis, differences between the processing times of the SMC and MCMC approaches were also very relevant: our proposal updated the marginal posterior distribution of the parameters, hyperparameters, and random effects in approximately 193 minutes (without activating the *resampling* and *moving* steps this time was of 16 minutes), while MCMC methods via `JAGS` software spent about 867 minutes.

To conclude, we have learnt that the combination of the Bayesian joint models for longitudinal and time-to-event data and the sequential methodology is an extremely useful and powerful tool. In particular, our results have showed that combining them is a good alternative to reduce computational time in joint models framework with sequential data configurations. Finally, it is worth mentioning that our proposal is, to our knowledge, the first one that fully integrates the Bayesian joint modelling and sequential methods.

6.2   Future work

Bayesian joint modelling of longitudinal and time-to-event data is a very relevant subject, both applied and methodological. Consequently, a great number of potentially interesting areas of research

can be considered.

As for the Bayesian joint models, other modelling can take advantage of the sequential approach. For instance, more complex joint models in relation to the number of longitudinal biomarkers and/or spatial components as well as purely time-to-event analysis (e.g. using the well-known accelerated failure time or cure rate models), in which the sequential Monte Carlo approach has not yet been explored. However, the most interesting and innovative challenge would be to develop sequential update procedures for non-static joint models. Another interesting scenario is joint models with non-linear effects.

In the sequential Monte Carlo framework, an imminent difficulty is to define in advance the number of particles that a specific application will require. Hence, it is also relevant to develop automatic mechanisms for this choice. Perhaps, some computational improvement can be made from other integration method proposals (e.g. randomised quasi-Monte Carlo) to obtain the marginal likelihood function of the parameters and hyperparameters of the joint model. A modification that would surely reduce computational time is the implementation in parallel, since the independence between the particles makes this feasible. Running times can also be improved by adopting a different MCMC kernel in the *moving* step. It would also be interesting to analyse the performance of the sequential approach for joint modelling with a small (or big) data set, because in these cases the accuracy and the computational time may play an even more relevant role.

Finally, our main purpose is to develop an `R` package which supports
different structures of joint models for longitudinal and time-to-
event data, and allows to use the sequential methodology proposed
in this thesis.

Another possible direction for future research are the determinis-
tic approaches, such as variational Bayesian methods (Beal, 2003)
and integrated nested Laplace approximations (Rue *et al.*, 2009),
to approximate marginal posterior distributions. Both methodolo-
gies work very fast in many applications, including some sequential
scenarios.

JAGS code

Joint model (4.2) - Simulated study

```
1  model{
2
3   for(i in 1:N){
4       for(j in 1:n[i]){
5           # LONGITUDINAL PROCESS
6           y[i,j] ~ dnorm(mu[i,j],tau)
7           mu[i,j] <- beta0+b0[i]+(beta1+b1[i])*time[i,j]
8
9           # TIME-TO-DROPOUT PROCESS
10          delta[i,j] ~ dbern(p[i,j])
11          logit(p[i,j]) <- gamma + alpha*mu[i,j]
12      }
13
14      # RANDOM EFFECTS
15      b0[i] ~ dnorm(0,tau0)
16      b1[i] ~ dnorm(0,tau1)
17  }
18
19  # MARGINAL PRIOR DISTRIBUTIONS
20  beta0 ~ dnorm(0,0.01)
21  beta1 ~ dnorm(0,0.01)
22  tau <- pow(sigma,-2)
23  sigma ~ dunif(0,10)
24  tau0 <- pow(sigma0,-2)
25  sigma0 ~ dunif(0,10)
26  tau1 <- pow(sigma1,-2)
27  sigma1 ~ dunif(0,10)
28  gamma ~ dnorm(0,0.01)
29  alpha ~ dnorm(0,0.01)
30  }
```

Joint model (5.1) and (5.3) - Application

```
 1  model{
 2
 3   for(i in 1:N){
 4       # LONGITUDINAL PROCESS
 5       for(j in 1:n[i]){
 6           y[i,j] ~ dnorm(mu[i,j],tau)
 7           mu[i,j] <- beta[1]+b[i,1]+(beta[2]+b[i,2])*day[i,j]+beta[3]*age[i]
 8       }
 9
10       # COMPETING RISKS PROCESS
11       # Calculation for integration (cumulative hazard)
12       for(j in 1:K){
13           # Hazard function for alive
14           h1[i,j] <- nu1*pow(Time[i]/2*(xk[j]+1),nu1-1)*exp(lambda1+
15               gamma1*age[i]+alpha01*b[i,1]+alpha11*b[i,2]*(Time[i]/2*(xk[j]+1)))
16           # Hazard function for dead
17           h2[i,j] <- nu2*pow(Time[i]/2*(xk[j]+1),nu2-1)*exp(lambda2+
18               gamma2*age[i]+alpha02*b[i,1]+alpha12*b[i,2]*(Time[i]/2*(xk[j]+1)))
19       }
20       # Cumulative hazard H[t] = int_0^t h[u] du - Gauss-Legendre quadrature
21       cumHaz1[i] <- (Time[i]/2)*inprod(wk,h1[i,])
22       cumHaz2[i] <- (Time[i]/2)*inprod(wk,h2[i,])
23       lS1[i] <- - cumHaz1[i]
24       lS2[i] <- - cumHaz2[i]
25
26       # RANDOM EFFECTS
27       b0[i] ~ dnorm(0,tau0)
28       b1[i] ~ dnorm(0,tau1)
29
30       # Definition of the log-likelihood using zeros trick
31       phi[i] <- C-(delta1[i]*log(h1[i,K])+delta2[i]*log(h2[i,K])+lS1[i]+lS2[i])
32       zeros[i] ~ dpois(phi[i])
33   }
34
35   # MARGINAL PRIOR DISTRIBUTIONS
36   for(i in 1:3){beta[i] ~ dnorm(0,0.001)}
37   tau <- pow(sig,-2)
38   sigma ~ dunif(0,100)
39   tau0 <- pow(sig0,-2)
40   sigma0 ~ dunif(0,50)
41   tau1 <- pow(sig1,-2)
42   sigma1 ~ dunif(0,50)
43   gamma1 ~ dnorm(0,0.001)
44   alpha01 ~ dnorm(0,0.001)
45   alpha11 ~ dnorm(0,0.001)
46   nu1 ~ dgamma(0.01,0.01)
47   lambda1 ~ dnorm(0,0.001)
48   gamma2 ~ dnorm(0,0.001)
49   alpha02 ~ dnorm(0,0.001)
50   alpha12 ~ dnorm(0,0.001)
51   nu2 ~ dgamma(0.01,0.01)
52   lambda2 ~ dnorm(0,0.001)
53  }
```

Usual probability distributions

Bernoulli distribution

A random variable X has a Bernoulli distribution $X \sim \mathcal{B}(p)$ with parameter $0 < p < 1$ if its probability mass function is described by

$$f(x \mid p) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases} \quad \text{for } x \in \{0, 1\}.$$

Its mean and variance are

$$\mathrm{E}(X) = p \quad \text{and} \quad \mathrm{Var}(X) = p(1 - p).$$

Gamma distribution

A random variable X has a gamma distribution $X \sim \mathcal{G}(\alpha, \beta)$ with parameters $\alpha > 0$ and $\beta > 0$ if its probability density function is described by

$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}, \quad \text{for } x \geq 0.$$

Its mean and variance are

$$E(X) = \frac{\alpha}{\beta} \quad \text{and} \quad \text{Var}(X) = \frac{\alpha}{\beta^2}.$$

Normal distribution

A random variable X has a normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu \in \Re$ and variance $\sigma^2 > 0$ if its probability density function is described by

$$f(x \mid \mu, \sigma) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right], \quad \text{for } x \in \Re.$$

Its mean and variance are

$$E(X) = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2.$$

Uniform distribution

A random variable X has a (continuous) uniform distribution $X \sim \mathcal{U}(a, b)$ with parameters $-\infty < a < b < \infty$ if its probability density function is described by

$$f(x \mid a, b) = \frac{1}{b - a}, \quad \text{for } a \leq x \leq b.$$

Its mean and variance are

$$E(X) = \frac{a + b}{2} \quad \text{and} \quad \text{Var}(X) = \frac{(b - a)^2}{12}.$$

Weibull distribution

A random variable X has a Weibull distribution $X \sim \mathcal{W}(\lambda, \nu)$ with parameters $\lambda > 0$ and $\nu > 0$ if its probability density function is described by

$$f(x \mid \lambda, \nu) = \lambda \nu x^{\nu-1} \exp(-\lambda x^{\nu}), \quad \text{for } x \geq 0.$$

Its mean and variance are

$$\mathrm{E}(X) = \frac{\Gamma(1 + 1/\nu)}{\lambda} \quad \text{and}$$

$$\mathrm{Var}(X) = \frac{1}{\lambda^2} \left\{ \Gamma\left(1 + \frac{2}{\nu}\right) - \left[\Gamma\left(1 + \frac{1}{\nu}\right)\right]^2 \right\}.$$

Simulation studies

In the tables below, we display the posterior summary for various simulated scenarios from Algorithm 3 presented in Chapter 4 using the joint model (4.2). Convergence in all these scenarios was achieved using the following MCMC configuration: three Markov chains with 200000 iterations after a burn-in period of 50000 iterations, and storing every 400th iteration. The marginal prior distributions were set as in (4.4) and, in the sequential procedure, the number of particles employed was 500 and the approximation of the marginal likelihood integrating out the random effects was performed by the quasi-Monte Carlo method with the Halton proposal (for more details, see Section 4.4). In each table, we have included the marginal posterior expectation and the standard deviation for the inferential results corresponding to the initial stage with the first data set and after incorporating the information of a new individual in the analysis. **SMC** represents the sequential update from the Algorithm 2 in Chapter 3 and **JAGS** the non-sequential update from `JAGS` software. Computational time from each approach

is also shown. In order to obtain a greater impact on the inferential update, we have artificially created an "outlier individual" with respect to the simulated scenario. In the description of each table we have included the longitudinal information of this new individual.

- Longitudinal information of the new individual incorporated in the analysis: $0, 5, 4, 6, 3, 6, 5, 7, 5, 4, 6$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	-1.0	-0.958	0.095	-0.904	0.102	-0.904	0.104
β_1	0.1	0.087	0.033	0.086	0.033	0.086	0.033
σ	1.5	1.483	0.041	1.491	0.043	1.490	0.043
σ_0	0.2	0.203	0.128	0.434	0.150	0.434	0.152
σ_1	0.2	0.221	0.023	0.231	0.023	0.231	0.025
γ	-2.5	-2.520	0.158	-2.584	0.158	-2.585	0.158
α	0.3	0.144	0.146	0.065	0.127	0.065	0.129
Time (minutes)		79		5		79	

- Longitudinal information of the new individual incorporated in the analysis: $0, -4, -3, -5, -2, -4, -3, -1, -3, -4, -3$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	0.0	-0.021	0.076	-0.047	0.082	-0.047	0.081
β_1	0.3	0.318	0.033	0.315	0.036	0.315	0.036
σ	1.0	0.998	0.034	1.005	0.035	1.007	0.035
σ_0	0.3	0.244	0.127	0.398	0.115	0.399	0.116
σ_1	0.2	0.192	0.029	0.205	0.030	0.205	0.031
γ	-2.0	-2.117	0.191	-2.158	0.181	-2.161	0.181
α	0.5	0.392	0.129	0.419	0.124	0.420	0.124
Time (minutes)		56		5		56	

- Longitudinal information of the new individual incorporated in the analysis: $0, 8, 10, 6, 7, 5, 8, 4, 5, 7, 6$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	1.0	1.111	0.137	1.162	0.143	1.163	0.142
β_1	-0.5	-0.567	0.058	-0.568	0.060	-0.568	0.061
σ	2.0	1.958	0.059	1.985	0.061	1.987	0.062
σ_0	0.1	0.218	0.157	0.386	0.213	0.386	0.213
σ_1	0.3	0.366	0.041	0.398	0.045	0.398	0.046
γ	-2.0	-2.191	0.149	-2.198	0.145	-2.196	0.147
α	-0.4	-0.404	0.074	-0.407	0.071	-0.407	0.072
Time (minutes)		67		5		67	

- Longitudinal information of the new individual incorporated in the analysis: $10, 7, 9, 5, 8, 11, 9, 10, 8, 8, 7$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	3.0	3.007	0.044	3.093	0.075	3.093	0.076
β_1	1.0	1.003	0.056	0.972	0.058	0.972	0.058
σ	0.5	0.512	0.020	0.581	0.022	0.580	0.022
σ_0	0.3	0.159	0.077	0.620	0.061	0.619	0.060
σ_1	0.5	0.521	0.043	0.505	0.043	0.508	0.043
γ	-5.0	-5.848	0.534	-5.383	0.524	-5.385	0.523
α	0.7	0.827	0.090	0.724	0.086	0.724	0.086
Time (minutes)		39		5		39	

- Longitudinal information of the new individual incorporated in the analysis: $5, 8, 6, 10, 6, 7, 5, 7, 6, 7, 6$.

θ	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	0.0	0.030	0.070	0.146	0.103	0.148	0.103
β_1	0.0	-0.005	0.044	-0.044	0.042	-0.045	0.044
σ	1.0	0.991	0.030	1.011	0.032	1.011	0.032
σ_0	0.1	0.144	0.099	0.317	0.086	0.317	0.087
σ_1	0.3	0.327	0.033	0.308	0.032	0.309	0.032
γ	-2.0	-2.155	0.153	-2.154	0.134	-2.155	0.135
α	0.6	0.981	0.201	0.730	0.087	0.730	0.087
Time (minutes)		57		5		57	

- Longitudinal information of the new individual incorporated in the analysis: $-5, -6, -3, -7, -9, -2, 0, -1, 0, -2, 2$.

θ	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	2.0	2.219	0.201	2.118	0.223	2.120	0.226
β_1	2.0	1.906	0.069	1.905	0.075	1.905	0.075
σ	3.0	3.022	0.084	3.015	0.086	3.016	0.088
σ_0	0.2	0.364	0.251	0.988	0.320	0.988	0.319
σ_1	0.4	0.429	0.047	0.493	0.055	0.494	0.055
γ	-4.0	-3.987	0.338	-4.006	0.336	-4.009	0.336
α	0.2	0.186	0.026	0.190	0.025	0.188	0.026
Time (minutes)		66		5		66	

- Longitudinal information of the new individual incorporated in the analysis: $-6, -5, -4, -6, -5, -3, -1, 0, -1, -2, 0$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	-1.0	-1.025	0.178	-1.075	0.190	-1.075	0.191
β_1	1.5	1.510	0.062	1.480	0.068	1.480	0.068
σ	2.5	2.594	0.083	2.587	0.082	2.587	0.082
σ_0	0.5	0.348	0.229	0.575	0.292	0.575	0.291
σ_1	0.3	0.313	0.048	0.380	0.052	0.381	0.052
γ	-1.5	-1.449	0.161	-1.513	0.148	-1.514	0.151
α	-0.1	-0.125	0.032	-0.118	0.033	-0.118	0.031
Time (minutes)		43		5		43	

- Longitudinal information of the new individual incorporated in the analysis: $3, 7, 5, 3, 4, 6, 8, 7, 7, 5, 8$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	-2.0	-2.058	0.281	-1.925	0.277	-1.927	0.278
β_1	-1.0	-0.980	0.103	-1.000	0.107	-1.002	0.107
σ	4.0	4.008	0.128	3.990	0.126	3.990	0.128
σ_0	0.4	0.357	0.255	0.483	0.328	0.483	0.328
σ_1	0.5	0.450	0.069	0.557	0.074	0.559	0.074
γ	-3.0	-3.047	0.338	-3.006	0.322	-3.005	0.322
α	-0.3	-0.257	0.055	-0.249	0.055	-0.249	0.053
Time (minutes)		53		5		53	

- Longitudinal information of the new individual incorporated in the analysis: $5, 10, 8, 8, 8, 7, 5, 6, 6, 2, 4$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	1.5	1.514	0.207	1.643	0.222	1.644	0.222
β_1	-2.0	-2.015	0.085	-2.031	0.091	-2.035	0.092
σ	3.0	2.986	0.091	2.980	0.094	2.981	0.093
σ_0	0.3	0.306	0.216	0.546	0.307	0.547	0.307
σ_1	0.5	0.493	0.057	0.586	0.064	0.586	0.062
γ	-5.0	-5.211	0.550	-5.132	0.552	-5.134	0.552
α	-0.6	-0.589	0.074	-0.574	0.073	-0.575	0.073
Time (minutes)		60		5		60	

- Longitudinal information of the new individual incorporated in the analysis: $0, 5, 4, 6, 7, 8, 7, 6, 3, 0, 1$.

$\boldsymbol{\theta}$	True	Initial		SMC		JAGS	
		Mean	SD	Mean	SD	Mean	SD
β_0	0.2	0.176	0.104	0.240	0.114	0.239	0.114
β_1	-2.0	-2.014	0.070	-2.001	0.075	-2.000	0.075
σ	1.5	1.517	0.042	1.564	0.050	1.564	0.047
σ_0	0.1	0.185	0.129	0.461	0.186	0.462	0.187
σ_1	0.6	0.623	0.050	0.668	0.056	0.670	0.056
γ	-6.0	-5.512	0.452	-5.540	0.458	-5.536	0.459
α	-0.5	-0.422	0.043	-0.425	0.043	-0.423	0.043
Time (minutes)		67		5		67	

Bibliography

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes.* Springer-Verlag, 1st edition, 1993. ISBN 9781461243489.

E. R. Andrinopoulou, D. Rizopoulos, M. L. Geleijnse, E. Lesaffre, A. J. J. C. Bogers, and J. J. Takkenberg. Dynamic prediction of outcome for patients with severe aortic stenosis: application of joint models for longitudinal and time-to-event data. *BMC Cardiovascular Disorders*, 15(28):1 – 8, 2015a.

E. R. Andrinopoulou, D. Rizopoulos, J. J. Takkenberg, and E. Lesaffre. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research*, 0(0):1 – 18, 2015b.

C. Armero, C. Forné, M. Rué, A. Forte, H. Perpiñán G. Gómez, and M. Baré. Bayesian joint ordinal and survival modeling for breast cancer risk assessment. *Statistics in Medicine*, 35(28):5267 – 5282, 2016a.

C. Armero, A. Forte, H. Perpiñán, M. J. Sanahuja, and S. Agustí. Bayesian joint modeling for assessing the progression of chronic kidney disease in children. *Statistical Methods in Medical Research*, 0(0):1 – 17, 2016b.

M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174 – 188, 2002.

O. Asar, J. Ritchie, P. A. Kalra, and P. J. Diggle. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International Journal of Epidemiology*, 44(1):334 – 344, 2015.

T. Baghfalaki, M. Ganjali, and R. Hashemi. Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics*, 24(4):834 – 855, 2014.

S. Balakrishnan and D. Madigan. A one-pass sequential Monte Carlo method for Bayesian analysis of massive datasets. *Bayesian Analysis*, 1(2):345 – 362, 2006.

S. Bandyopadhyay, B. Ganguli, and A. Chatterjee. A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research*, 20(4):299 – 330, 2011.

D. Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 1st edition, 2012. ISBN 9780521518147.

J. Barrett and L. Su. Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in Medicine*, 36(9): 1447 – 1460, 2017.

M. J. Bayarri and J. O. Berger. The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19(1):58 – 80, 2004.

M. J. Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London, 2003.

J. M. Bernardo and A. F. M. Smith. *Bayesian theory*. John Wiley & Sons, 1st edition, 1994. ISBN 9780471924166.

C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza. Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association*, 92(440):1403 – 1412, 1997.

V. Bewick, L. Cheek, and J. Ball. Statistics review 12: survival analysis. *Critical Care*, 8(5):389 – 394, 2004.

E. Bonawitz, S. Denison, A. Gopnik, and T. L. Griffiths. Win-stay, Lose-sample: a simple sequential algorithm for approximating Bayesian inference. *Conginitive Psychology*, 74:35 – 65, 2014.

J. M. Box-Steffensmeier and B. S. Jones. *Event history modeling: a guide for social scientists.* Cambridge University Press, 1st edition, 2004. ISBN 9780521546737.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434 – 455, 1998.

E. R. Brown and J. G. Ibrahim. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2): 221 – 228, 2003.

O. Cappé, E. Moulines, and T. Ryden. *Inference in hidden Markov models.* Springer-Verlag, 1st edition, 2005. ISBN 9780387402642.

O. Cappé, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. In *Proceedings of the Institute of Electrical and Electronics Engineers*, volume 95(5), pages 899 – 924, 2007.

O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. Robert. Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18(4):447 – 459, 2008.

B. P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(3):473 – 484, 1995.

J. B. Carlin, R. Wolfe, C. H. Brown, and A. Gelman. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics*, 2(4):397 – 416, 2001.

J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for non-linear problems. In *IEE Proceedings Radar, Sonar and Navigation*, volume 146(1), pages 2 – 7, 1999.

B. Chakra-Borty and E. E. M. Moodie. *Statistical methods for dynamic treatment regimes: reinforcement learning, causal inference, and personalized medicine.* Springer-Verlag, 1st edition, 2013. ISBN 9781461474272.

J. S. K. Chan. Bayesian informative dropout model for longitudinal binary data with random effects using conditional and joint modeling approaches. *Biometrical Journal*, 58(3):549 – 569, 2016.

S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327 – 335, 1995.

S. H. Cho, J. Jeon, and S. Kim. Personalized medicine in breast cancer: a systematic review. *Journal of Breast Cancer*, 15(3):265 – 272, 2012.

N. Chopin. A sequential particle filter for static models. *Biometrika*, 89 (3):539 – 551, 2002.

N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *The Annals of Statistics*, 32 (6):2385 – 2411, 2004.

N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 75(3):397 – 426, 2013.

B. A. Clegg, G. J. DiGirolamo, and S. W. Keele. Sequence learning. *Trends in Cognitive Sciences*, 2(8):275 – 281, 1998.

D. Collett. *Modelling survival data in medical research*. Chapman & Hall/CRC, 2nd edition, 2003. ISBN 9781584883258.

L. Collette, J. Bogaerts, S. Suciu, C. Fortpied, T. Gorlia, C. Coens, M. Mauer, B. Hasan, S. Collette, M. Ouali, S. Litière, J. Rapion, and R. Sylvester. Statistical methodology for personalized medicine: new developments at EORTC Headquarters since the turn of the 21st Century. *European Journal of Cancer Supplements*, 10(1):13 – 19, 2012.

E. A. Cornea, L. M. Chen, B. F. Qaqish, H. Chu, and J. G. Ibrahim. *Package JMdesign: joint modeling of longitudinal and survival data - power calculation*, 2014.

V. Couallier, L. Gerville-Reache, C. Huber-Carol, N. Limnios, and M. Mesbah, editors. *Statistical models and methods for reliability and survival analysis*. John Wiley & Sons, 1st edition, 2013. ISBN 9781848216198.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187 – 220, 1972.

D. Creal. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, 31(3):245 – 296, 2012.

D. Crisan and B. Rozovskii, editors. *The Oxford handbook of non-linear filtering.* Oxford University Press, 1st edition, 2011. ISBN 9780199532902.

M. J. Crowther, K. R. Abrams, and P. C. Lambert. Joint modeling of longitudinal and survival data. *Stata Journal*, 13(1):165 – 184, 2013.

M. J. Daniels and J. W. Hogan. *Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis.* Chapman & Hall/CRC, 1st edition, 2008. ISBN 9781584886099.

V. De Gruttola and X. M. Tu. Modeling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50(4):1003 – 1014, 1994.

P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Methodological)*, 68 (3):411 – 436, 2006.

P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli*, 18(1):252 – 278, 2012.

L. Devroye. *Non-uniform random variate generation.* Springer-Verlag, 1st edition, 1986. ISBN 9781461386452.

T. G. Dietterich. *Structural, syntactic, and statistical pattern recognition*, chapter Machine learning for sequential data: a review, pages 15 – 30. Springer Berlin Heidelberg, 1st edition, 2002. ISBN 9783540440116.

P. J. Diggle, P. Heagerty, K. Y. Liang, and S. L. Zeger. *Analysis of longitudinal data.* Oxford University Press, 2nd edition, 2002. ISBN 9780199676750.

R. Douc and O. Cappé. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64 – 69, 2005.

R. Douc, E. Moulines, and J. Olsson. Long-term stability of sequential Monte Carlo methods under verifiable conditions. *The Annals of Applied Probability*, 24(5):1767 – 1802, 2014.

A. Doucet and A. M. Johansen. *The Oxford handbook of nonlinear filtering*, chapter A tutorial on particle filtering and smoothing: fifteen years later, pages 656 – 704. Oxford University Press, 1st edition, 2011. ISBN 9780199532902.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3): 197 – 208, 2000.

A. Doucet, N. Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Springer-Verlag, 1st edition, 2001. ISBN 9781441928870.

C. C. Drovandi, J. M. McGree, and A. N. Pettitt. Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data. *Computational Statistics and Data Analysis*, 57(1):320 – 335, 2013.

C. C. Drovandi, J. M. McGree, and A. N. Pettitt. A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1): 3 – 24, 2014.

R. M. Elashoff, G. Li, and N. Li. *Joint modeling of longitudinal and time-to-event data.* Chapman and Hall/CRC, 1st edition, 2016. ISBN 9781439807828.

T. Emura. *Package joint.Cox: penalized likelihood estimation and dynamic prediction under the joint frailty-copula models between tumour progression and death for meta-analysis*, 2016.

Y. Fan, D. S. Leslie, and M. P. Wand. Generalised linear mixed model analysis via sequential Monte Carlo sampling. *Electronic Journal of Statistics*, 2:916 – 938, 2008.

C. L. Faucett and D. C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, 15(15):1663 – 1685, 1996.

H. Faure. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arithmetica*, 41(4):337 – 351, 1982.

P. Fearnhead and B. M. Taylor. An adaptive sequential Monte Carlo sampler. *Bayesian Analysis*, 8(2):411 – 438, 2013.

G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, editors. *Longitudinal data analysis.* Chapman & Hall/CRC, 1st edition, 2008. ISBN 9781584886587.

US Food and Drug Administration. *Paving the way for personalized medicine: FDA's role in a new era of medical product development.* Create Space Independent Publishing Platform, 1st edition, 2014. ISBN 9781496083203.

N. Freitas, M. Milo, P. Clarkson, M. Niranjan, and A. Gee. Sequential support vector machines. In *Proceedings of the 1999 IEEE Signal Processing Society Workshop*, volume 1, pages 31 – 40, 1999.

A. Galecki and T. Burzykowski. *Linear mixed-effects models using R: a step-by-step approach.* Springer New, 1st edition, 2013. ISBN 9781461438991.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference.* Chapman & Hall/CRC, 2nd edition, 2006. ISBN 9781584885870.

M. Ganjali and T. Baghfalaki. A copula approach to joint modeling of longitudinal measurements and survival times using Monte Carlo expectation-maximization with application to AIDS studies. *Journal of Biopharmaceutical Statistics*, 25(5):1077 – 1099, 2015.

M. Gao and H. Zhang. Sequential Monte Carlo methods for parameter estimation in nonlinear state-space models. *Computers & Geosciences*, 44:70 – 77, 2012.

A. E. Gelfand. Gibbs sampling. *Journal of the American Statistical Association*, 95(452):1300 – 1304, 2000.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398 – 409, 1990.

A. Gelman and J. Hill. *Data analysis using regression and multi-level/hierarchical models.* Cambridge University Press, 1st edition, 2006. ISBN 9780521686891.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457 – 472, 1992.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721 – 741, 1984.

M. Gerber and N. Chopin. Sequential quasi Monte Carlo. *Journal of the Royal Statistical Society: Series B (Methodological)*, 77(3):509 – 579, 2015.

W. R. Gilks and C. Berzuini. Following a moving target: Monte Carlo inference for dynamic Bayesian model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 63(1):127 – 146, 2001.

W. R. Gilks, D. G. Clayton, D. J. Spiegelhalter, N. G. Best, A. J. McNeil, L. D. Sharples, and A. J. Kirby. Modelling complexity: applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):39 – 52, 1993.

M. E. Glickman and D. A. van Dyk. *Topics in biostatistics*, chapter Basic Bayesian methods, pages 319 – 338. The Humana Press Inc., 1st edition, 2007. ISBN 9781597455305.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140(2), pages 107 – 113, 1993.

A. L. Gould, M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef, and F. Y. Bois. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34 (14):2181 – 2195, 2014.

X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):1 – 9, 2004.

J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84 – 90, 1960.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer-Verlag, 2nd edition, 2009. ISBN 9780387848570.

R. Henderson, P. J. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(44):465 – 480, 2000.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593 – 1623, 2014.

J. W. Hogan and N. M. Laird. Mixture models for the joint distributions of repeated measures and event times. *Statistics in Medicine*, 16(3): 239 – 257, 1997a.

J. W. Hogan and N. M. Laird. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, 16(3):259 – 272, 1997b.

J. W. Hogan and N. M. Laird. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research*, 7(1):28 – 48, 1998.

J. D. Hol, T. B. Schon, and F Gustafsson. On resampling algorithms for particle filters. In *Proceedings of the Nonlinear Statistical Signal Processing Workshop*, pages 79 – 82, 2006.

W. Hu, G. Li, and N. Li. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*, 28(11):1601 – 1619, 2009.

X. Huang, G. Li, and R. M. Elashoff. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and Its Interface*, 3(2): 185 – 195, 2010.

Y. Huang, G. Dagne, and L. Wu. Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine*, 30(24):2930 – 2946, 2011.

Y. Huang, X. J. Hu, and G. A. Dagne. Jointly modeling time-to-event and longitudinal data: a Bayesian approach. *Statistical Methods and Applications*, 23(1):95 – 121, 2014.

J. G. Ibrahim and G. Molenberghs. Missing data methods in longitudinal studies: a review. *TEST*, 18(1):1 – 43, 2009.

J. G. Ibrahim, M. H. Chen, and D. Sinha. *Bayesian survival analysis.* Springer-Verlag, 1st edition, 2001. ISBN 9780387989358.

J. G. Ibrahim, M. H. Chen, and D. Sinha. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, 14:863 – 883, 2004.

J. G. Ibrahim, H. Chu, and L. M. Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796 – 2801, 2010.

K. K. Jain. *Textbook of personalized medicine.* Humana Press, 1st edition, 2015. ISBN 9781493925520.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning with applications in R.* Springer New York, 1st edition, 2013. ISBN 9781461471370.

N. P. Jewell, A. C. Kimber, M. L. T. Lee, and G. A. Whitmore, editors. *Lifetime data: models in reliability and survival analysis.* Springer US, 1st edition, 1996. ISBN 9780792337836.

M. Jouin, R. Gouriveau, D. Hissel, M. C. Péra, and N. Zerhouni. Particle filter-based prognostics: review, discussion and perspectives. *Mechanical Systems and Signal Processing*, 72/73:2 – 31, 2016.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data.* John Wiley & Sons, 2nd edition, 2002. ISBN 9780471363576.

N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski. An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In *Proceedings of the 15th IFAC Symposium on System Identification*, volume 42(10), pages 774 – 785, 2009.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773 – 795, 1995.

R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal. Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52(2):93 – 100, 1998.

H. T. Kim. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research*, 13(2):559 – 565, 2007.

G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian non-linear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1 – 25, 1996.

J. P. Klein and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer-Verlag, 2nd edition, 2003. ISBN 9780387216454.

D. Kleinbaum and M. Klein. *Survival analysis: a self-learning text*. Springer-Verlag, 3th edition, 2012. ISBN 9781441966452.

N. Laird and J. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963 – 974, 1982.

E. T. Lee and J. W. Wang. *Statistical methods for survival data analysis*. John Wiley & Sons, 4th edition, 2013. ISBN 9781118095027.

J. J. Lee and C. T. Chu. Bayesian clinical trials in action. *Statistics in Medicine*, 31(25):2955 – 2972, 2012.

K. M. Leung, R. M. Elashoff, and A. A. Afifi. Censoring issues in survival analysis. *Annual Review of Public Health*, 18:83 – 104, 1997.

R. Little. *Longitudinal data analysis*, chapter Selection and pattern-mixture models, pages 409 – 431. Chapman & Hall/CRC, 1st edition, 2008. ISBN 9781584886587.

R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2nd edition, 2002. ISBN 9780471183860.

J. S. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443): 1032 – 1044, 1998.

H. F. Lopes and R. S. Tsay. Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168 – 209, 2011.

T. J. Loredo. *Maximum entropy and Bayesian methods*, chapter From Laplace to supernova SN 1987A: Bayesian inference in astrophysics, pages 81 – 142. Springer Netherlands, 1st edition, 1990. ISBN 9789401067928.

T. J. Loredo. *Statistical challenges in modern astronomy*, chapter The promise of Bayesian inference for astrophysics, pages 275 – 297. Springer, 1st edition, 1992. ISBN 9781461392927.

Y. Lu, J. Fang, L. Tian, and H. Jin, editors. *Advanced medical statistics*. World Scientific Publishing Company, 2nd edition, 2015. ISBN 9789814583299.

D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325 – 337, 2000.

R. Martins, G. L. Silva, and V. Andreozzi. Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Statistics in Medicine*, 35(19):3368 – 3384, 2016.

A. Mauguen, B. Rachet, S. Mathoulin-Pélissier, G. MacGrogan, A. Laurent, and V. Rondeau. Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine*, 32(30):5366 – 5380, 2013.

P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman & Hall/CRC, 2nd edition, 1989. ISBN 9780412317606.

S. B. McGrayne. *The theory that would not die*. Yale University Press, 1st edition, 2011. ISBN 9780300169690.

J. M. McGree, C. C. Drovandi, G. White, and A. N. Pettitt. A pseudo-marginal sequential Monte Carlo algorithm for random effects models in Bayesian sequential design. *Statistics and Computing*, 26(5):1121 – 1136, 2016.

J. L. Mega, M. S. Sabatine, and E. M. Antman. Population and personalized medicine in the modern era. *The Journal of the American Medical Association*, 312(19):1969 – 1970, 2014.

R. M. Merrill. *Statistical methods in epidemiologic research*. Jones & Bartlett Learning, 1st edition, 2015. ISBN 9781284050202.

M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. The MIT Press, 1st edition, 2012. ISBN 9780262018258.

G. Molenberghs and M. Kenward. *Missing data in clinical studies*. John Wiley & Sons, 1st edition, 2007. ISBN 9780470849811.

G. Molenberghs, G. Fitzmaurice, M. G. Kenward, A. Tsiatis, and G. Verbeke, editors. *Handbook of missing data methodology*. Chapman & Hall/CRC, 1st edition, 2014. ISBN 9781439854617.

C. Montzka, V. R. N. Pauwels, H. J. H. Franssen, X. Han, and H. Vereecken. Multivariate and multiscale data assimilation in terrestrial systems: a review. *Sensors*, 12(12):16291 – 16333, 2012.

W. J. Morokoff and R. E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251 – 1279, 1994.

R. Murugan. Movement towards personalised medicine in the ICU. *The Lancet Respiratory Medicine*, 3(1):10 – 12, 2015.

B. Muthén, T. Asparouhov, M. E. Boye, M. Hackshaw, and A. Naegeli. Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. Technical report, Muthén & Muthén, 2009.

A. Neuhaus, T. Augustin, C. Heumann, and D. Daumer. A review on joint models in biometrical research. *Journal of Statistical Theory and Practice*, 3(4):855 – 868, 2009.

H. Niederreiter. Some current issues in quasi-Monte Carlo methods. *Journal of Complexity*, 19(3):428 – 433, 2003.

G. Petris, S. Petrone, and P. Campagnoli. *Dynamic linear models with R*. Springer-Verlag, 1st edition, 2009. ISBN 9780387772370.

P. Philipson, I. Sousa, P. J. Diggle, P. Williamson, R. Kolamunnage-Dona, and R. Henderson. *Package joineR: joint modelling of repeated measurements and time-to-event data*, 2012.

M. Pintilie. *Competing risks: a practical perspective*. John Wiley & Sons, 1st edition, 2006. ISBN 9780470870686.

M. Plummer. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003.

C. Proust-Lima and J. M. Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*, 10(3):535 – 549, 2009.

C. Proust-Lima, P. Joly, J. F. Dartigues, and H. Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics and Data Analysis*, 53(14):1142 – 1154, 2009.

C. Proust-Lima, V. Philipps, A. Diakite, and B. Liquet. *Package lcmm: extended mixed models using latent classes and latent processes*, 2010.

H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11): 2389 – 2430, 2007.

G. Ridgeway and D. Madigan. A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery*, 7(3):301 – 319, 2003.

D. Rizopoulos. JM: an R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1 – 33, 2010.

D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819 – 829, 2011.

D. Rizopoulos. *Package JMbayes: joint modeling of longitudinal and time-to-event data under a Bayesian approach*, 2012a.

D. Rizopoulos. *Joint models for longitudinal and time-to-event data: with applications in R.* Chapman and Hall/CRC, 1st edition, 2012b. ISBN 9781439872864.

D. Rizopoulos, G. Verbeke, and G. Molenberghs. Multiple-imputation-based residuals and diagnostic plot for joint models of longitudinal and survival outcomes. *Biometrics*, 66(1):20 – 29, 2010.

D. Rizopoulos, L. A. Hatfield, B. P. Carlin, and J. J. Takkenberg. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association*, 109(508):1385 – 1397, 2014.

C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation.* Springer New York, 2nd edition, 2007. ISBN 9780387715988.

C. Robert. Bayesian computational tools. *Annual Review of Statistics and Its Application*, 1(1):153 – 177, 2014.

C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer New York, 2nd edition, 2004. ISBN 9781441919397.

C. Robert and G. Casella. A short history of Markov chain Monte Carlo: subjective recollections from incomplete data. *Statistical Science*, 26 (1):102 – 115, 2011.

V. Rondeau, J. R. Gonzalez, Y. Mazroui, A. Mauguen, A. Krol, A. Diakite, A. Laurent, and M. Lopez. *Package frailtypack: general frailty models: shared, joint and nested frailty models with prediction*, 2017.

D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581 – 592, 1976.

D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm. *Journal of American Statistical Association*, 82(398):543 – 546, 1987.

D. B. Rubin. *Bayesian Statistics 3*, chapter Using the SIR algorithm to simulate posterior distribution, pages 395 – 402. Oxford University Press, 1st edition, 1988. ISBN 9780198522201.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 71(2):319 – 392, 2009.

M. Rué, E. R. Andrinopoulou, D. Alvares, C. Armero, A. Forte, and L. Blanch. Bayesian joint modeling of bivariate longitudinal and competing risks data: an application to study patient-ventilator asynchronies in critical care patients. *Biometrical Journal*, 2017. In press.

E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1):128 – 154, 2016.

A. Satty and H. Mwambi. Selection and pattern mixture models for modelling longitudinal data with dropout: an application study. *Statistics and Operations Research Transactions*, 37(2):131 – 152, 2013.

C. Schäfer and N. Chopin. Sequential Monte Carlo on large binary sampling spaces. *Statistics and Computing*, 23(2):163 – 184, 2013.

M. Schiefsky. *Hippocrates on ancient medicine*. Brill Academic Publishers, 1st edition, 2005. ISBN 9789004137585.

A. Sharratt. Personalized medicine: the future of benefits plans? *Benefits Canada*, 1:10 – 14, 2015.

I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86 – 112, 1967.

I. Sousa. A review on joint modeling of longitudinal measurements and time-to-event. *REVSTAT - Statistical Journal*, 9(1):57 – 81, 2011.

Z. Syed, P. Indyk, and J. Guttag. Learning approximate sequential patterns for classification. *Journal of Machine Learning Research*, 10:1913 – 1936, 2009.

L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701 – 1762, 1994.

C. Tseng, R. Elashoff, N. Li, and G. Li. Longitudinal data analysis with non-ignorable missing data. *Statistical Methods in Medical Research*, 25(1):205 – 220, 2016.

A. A. Tsiatis and M. Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809 – 834, 2004.

A. A. Tsiatis, V. De Gruttola, and M. S. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429):27 – 37, 1995.

G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data*. Springer-Verlag, 1st edition, 2000. ISBN 9780387950273.

G. Verbeke, B. Spiessens, and E. Lesaffre. Conditional linear mixed models. *The American Statistician*, 55(1):25 – 34, 2001.

G. Verbeke, S. Fieuws, G. Molenberghs, and M. Davidian. The analysis of multivariate longitudinal data: a review. *Statistical Methods in Medical Research*, 23(1):42 – 59, 2014.

E. F. Vonesh, T. Greene, and M. D. Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*, 25(1):143 – 163, 2006.

P. Wang, W. Shen, and M. E. Boye. Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. *Health Services and Outcomes Research Methodology*, 12(2/3):182 – 199, 2012.

R. E. Weiss. *Modeling longitudinal data*. Springer-Verlag, 1st edition, 2005. ISBN 9781441923219.

D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4 (2):65 – 85, 1994.

L. Wu, W. Liu, G. Y. Yi, and Y. Huang. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012:1 – 17, 2012.

M. C. Wu and K. Bailey. Analyzing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7(1/2):337 – 346, 1988.

M. C. Wu and R. J. Carrol. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring problem. *Biometrics*, 44:175 – 188, 1988.

M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330 – 339, 1997.

C. Xu, P. Z. Hadjipantelis, and J. L. Wang. *Package JSM: semiparametric joint modeling of survival and longitudinal data*, 2016.

W. Ye, X. Lin, and J. M. Taylor. A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface*, 1(1):35 – 45, 2008.

M. Yu, N. J. Law, J. M. Taylor, and H. M. Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14(3):835 – 862, 2004.

M. Yu, J. M. Taylor, and H. M. Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103(481):178 – 187, 2008.

D. Zhang, M. H. Chen, J. G. Ibrahim, M. E. Boye, and W. Shen. JMFit: a SAS macro for joint models of longitudinal and survival data. *Journal of Statistical Software*, 71(3):1 – 24, 2016.

M. J. Zhang, X. Zhang, and T. H. Scheike. Modeling cumulative incidence function for competing risks data. *Expert Review of Clinical Pharmacology*, 1(3):391 – 400, 2008.

Y. Zhao and D. Zeng. Recent development on statistical methods for personalized medicine discovery. *Frontiers of Medicine*, 7(1):102 – 110, 2013.

H. Zhu, J. G. Ibrahim, Y. Y. Chi, and N. Tang. Bayesian influence measures for joint models for longitudinal and survival data. *Biometrics*, 68(3):954 – 964, 2012.