Latent diffusion models for event history analysis

Gareth O. Roberts and Laura M. Sangalli

University of Warwick and Politecnico di Milano

Abstract

We consider Bayesian hierarchical models for event history analysis, where the event times are modeled through an underlying diffusion process, which determines the hazard rate. We show how these models can be efficiently treated by means of Markov chain Monte Carlo techniques.

Keywords and phrases. Diffusion processes, survival analysis, parametrization of hierarchical models.

1 Introduction

Diffusion processes have found many applications in the modelling of continuous-time phenomena, for problems related to several scientific areas, ranging from economics to biology, from physics to engineering. Here we use diffusion processes as building blocks for the definition of models for event history analysis. This idea is not new (see for example the reviews in Aalen and Gjessing (2001, 2004)). However, in this paper we are able to considerably extend the flexibility of the diffusion models used, by adopting powerful Markov Chain Monte Carlo techniques.

Diffusion models for survival analysis have been proposed because, as summarized in Aalen and Gjessing (2004), "when modelling survival data it may be of interest to imagine an underlying process leading up to the event in question". Such a process might for example represent the development of a disease. Two types of models have been considered in the literature. Models where the event happens when a diffusion process hits some barrier, and models where the hazard rate is some suitable function of the diffusion. For the former type of models, we refer the reader to Aalen and Gjessing (2001), and references therein. Here we are interested in the latter. Woodbury and Manton (1977) proposed a model where the hazard rate is a quadratic function of an Ornstein-Uhlenbeck diffusion process. This model has been later considered by several authors, including Myers (1981), Yashin (1985), Yashin and Vaupel (1986), and Aalen and Gjessing (2004). For given values of the parameters of the Ornstein-Uhlenbeck process, survival distributions and hazards are studied. Myers (1981) focuses on survival distributions conditioned on initial covariates values; Yashin (1985) and Yashin and Vaupel (1986) use hazards based on quadratic functions of Ornstein-Uhlenbeck processes in order to model heterogeneity among groups and among individuals, and study the relative hazard functions and survival distributions; Aalen and Gjessing (2004) derive quasi-stationary distributions. Obtaining such analytical results for hazard functions other than quadratic functions, or for more complex diffusion processes, is not feasible.

In our paper, we adopt a Bayesian approach and we show how these models can be efficiently

treated by means of Markov chain Monte Carlo techniques, for general choices of diffusion processes and hazard functions. We also consider the case of multiple groups of observations, typical of clinical trials. We test our MCMC algorithm both on simulated and on real data.

The paper is organized as follows. In Section 2 we recall the essential of diffusion processes and introduce the model. In Section 3 we describe the MCMC scheme. In Section 4 we discuss a straightforward generalization of the framework developed in the previous sections, and deal with the case of multiple groups of observations. Section 5 is devoted to simulation studies. In Section 6 we illustrate an improved version of the algorithm, based on a reparametrization of the model. In Section 7 we apply our models to a dataset from a clinical trial, that has been considered in a number of papers in the context of survival analysis, the famous Cox (1972) paper among the firsts. Finally, in Section 8 and 9 we discuss possible extensions of the model considered.

2 Latent diffusion models

Let Θ be a random variable with values in $\mathbb{R}^d$. Denote by $C([0,\infty),\mathbb{R})$ the space of continuous functions from $[0,\infty)$ to $\mathbb{R}$, and by $\mathcal{C}$ its cylinder σ-algebra. Given $\Theta = \theta$, consider the scalar diffusion process $X = \{X_t : t \geq 0\}$, solution of a *stochastic differential equation* (SDE, for short) of the form

$$\begin{aligned} dX_t &= \beta(X_t,\theta) + \sigma dB_t \qquad t \geq 0 \\ X_0 &= x_0 \end{aligned} \tag{1}$$

driven by the standard scalar Brownian motion $B = \{B_t : t \geq 0\}$. The Brownian motion B and the diffusion process X are random elements of $(C([0,\infty),\mathbb{R}),\mathcal{C})$. The diffusion coefficient σ is assumed constant and known, for the moment. The more technically difficult case of unknown σ is postponed to Section 8. The drift $\beta(x,\theta)$ is assumed to be jointly measurable in x and θ, and to satisfy the regularity conditions (locally Lipschitz, with linear growth bound) that guarantee the existence of a, weakly unique, global solution to (1). See, for example, Chapter V.24 in Rogers and Williams (2000).

Let $\mathbb{W}_\sigma$ be the law of σB, and, for a given θ, denote by $\mathbb{P}_\theta$ the law of the diffusion X, solution of (1). By *Girsanov's theorem*, the Radon-Nikodym derivative of $\mathbb{P}_\theta$, with respect to $\mathbb{W}_\sigma$, is given by

$$\frac{d\mathbb{P}_\theta}{d\mathbb{W}_\sigma}(x) = \exp\left\{\int_0^\infty \frac{\beta(x_t,\theta)}{\sigma^2}\,dx_t - \frac{1}{2}\int_0^\infty \frac{\beta(x_t,\theta)^2}{\sigma^2}\,dt\right\}.$$

See, for example, Chapter V.27 in Rogers and Williams (2000).

Similarly, for a finite T, denote by $C([0,T],\mathbb{R})$ the space of continuous functions from $[0,T]$ to $\mathbb{R}$, and by $\mathcal{C}^T$ its cylinder σ-algebra. Then, $B_{[0,T]} := \{B_t : 0 \leq t \leq T\}$ and $X_{[0,T]} = \{X_t : 0 \leq t \leq T\}$ are random elements of $(C([0,T],\mathbb{R}),\mathcal{C}^T)$. Let $\mathbb{W}_{T,\sigma}$ be the law of $\sigma B_{[0,T]}$, and, for a given θ, denote by $\mathbb{P}_{T,\theta}$ the law of $X_{[0,T]}$. Then, by Girsanov's theorem, the Radon-Nikodym derivative of $\mathbb{P}_{T,\theta}$, with respect to $\mathbb{W}_{T,\sigma}$, is given by

$$\frac{d\mathbb{P}_{T,\theta}}{d\mathbb{W}_{T,\sigma}}(x_{[0,T]}) = \exp\left\{\int_0^T \frac{\beta(x_t,\theta)}{\sigma^2}\,dx_t - \frac{1}{2}\int_0^T \frac{\beta(x_t,\theta)^2}{\sigma^2}\,dt\right\} \tag{2}$$

and, for each T, the measures $\mathbb{P}_{T,\theta}$ are absolutely continuous.

Given the diffusion X, let us consider the random distribution function $F_{X,h}$ on $[0,\infty)$, defined as

$$F_{X,h}(t) := 1 - \exp\left\{-\int_0^t h(X_s)\,ds\right\} \qquad t \geq 0 \tag{3}$$

where $h(\cdot)$ is some suitable nonnegative and continuous function, with $\int_0^\infty h(X_s)\,ds = \infty$ almost surely. The function $h(\cdot)$ plays the role of the hazard function, and $h(X_t)$ is the random hazard rate, at time t, associated to the random distribution $F_{X,h}$.

Two features of the random measure $F_{X,h}$ have to be noted. The first is that the hazard inherits the Markov property of the diffusion process, so that the hazard at a future time t' just depends on the hazard at the present time t. The Markov property seems indeed a sensible choice to make at the level of the hazard. The second is that the cumulative hazard is a process with positively correlated increments, being the integral of a continuous process. The latter feature is natural in many contexts, and it translates into the model the concern with the stochastic process that clearly must lie behind the occurrence of events. In words, an high increment of the cumulative hazard over the time interval $[t,t']$ means that the underlying stochastic process has reached a region of high risk, and this is likely to yield an high increment of the cumulative hazard over a close (disjoint) time interval. The strength of this positive correlation, and thus the smoothness of the cumulative hazard, depends on the choice of the hazard function h and of the diffusion process X: the rougher the diffusion, the weaker is the correlation, and viceversa. See also the comments in Section 9. Note that the property we have just highlighted differentiates the random distributions we are considering from another class of random distributions that has been extensively used in applications to event history analysis, namely the class of *neutral to the right random probabilities*. The cumulative hazards of these probabilities are processes with independent increments, and thus have an erratic behaviour. See Doksum (1974) for definition and properties of these random measures, and e.g. Kalbfleisch (1978), Hjort (1990) and Damien and Walker (2002) for applications in survival analysis. In fact, we could say that the random distribution $F_{X,h}$ is *positive to the right*: for each k and $0 < t_1 < t_2 < \cdots < t_k$, the normalised increments

$$F_{X,h}(t_1)\,, \quad \frac{F_{X,h}(t_2) - F_{X,h}(t_1)}{1 - F_{X,h}(t_1)}\,, \quad \cdots\,, \quad \frac{F_{X,h}(t_k) - F_{X,h}(t_{k-1})}{1 - F_{X,h}(t_{k-1})}$$

are positively correlated, instead of being stochastically independent as in the case of neutral to the right random probabilities.

Let us now consider a sequence of event times $Y_1, Y_2, \ldots$ which are, conditionally on $F_{X,h}$, independent and identically distributed (i.i.d., for short) with common distribution $F_{X,h}$. From (3), it follows that the distribution of $Y_1, \ldots, Y_n$, given $X = x$, has density, with respect to the n-dimensional Lebesgue measure $\mathcal{L}^n$, given by

$$l(y_1, \ldots, y_n | x) := \left[\prod_{j=1}^n h(x_{y_j})\right] \exp\left\{-\sum_{j=1}^n \int_0^{y_j} h(x_t)\,dt\right\}. \tag{4}$$

Censored observations can be easily dealt with in this setting. Suppose for example that the observations are censored if they exceed time C, then the likelihood becomes

$$l(y_1, \ldots, y_n | x) = \left[\prod_{j=1}^n h(x_{y_j})^{1(y_j<C)}\right] \exp\left\{-\sum_{j=1}^n \int_0^{y_j} h(x_t)\,dt\right\}.$$

We are thus considering a latent diffusion model for event history analysis, where the event times are modelled through an underlying diffusion process which determines the hazard rate. As highlighted by Aalen and Gjessing (2004), this model can be also interpreted as a random barrier hitting model. Indeed, the event happens when the cumulative hazard strike a random barrier R, which is exponentially distributed with mean 1, and is stochastically independent of X.

3 Markov Chain Monte Carlo methods for latent diffusion models

Let $p_\Theta(\theta)$ be the prior density, with respect to $\mathcal{L}^d$, of the d-dimensional parameter Θ, which appears in the drift of the diffusion process X, solution of (1). Fix a finite time horizon T of interest, with $T \geq y_{[n]}$, where $y_{[n]} := \max\{y_1, \ldots, y_n\}$. The choice of T will be discussed in Section 6. Then, the joint posterior distribution of Θ and $X_{[0,T]}$ has density, with respect to the product measure $\mathcal{L}^d \otimes \mathbb{W}_{T,\sigma}$, given by

$$\pi(\theta, x_{[0,T]}|y_1, \ldots, y_n) = C\, p_\Theta(\theta)\, g(x_{[0,T]}|\theta)\, l(y_1, \ldots, y_n|x_{[0,y_{[n]}]}) \tag{5}$$

where C is a normalizing constant, and $g(x_{[0,T]}|\theta) := \frac{d\mathbb{P}_{T,\theta}}{d\mathbb{W}_{T,\sigma}}(x)$ is given by Girsanov's formula (2).

A Gibbs sampling algorithm for sampling from (5) alternates between

1. simulation Θ, conditional on the observations and the current path of $X_{[0,T]}$;
2. simulation of $X_{[0,T]}$, conditional on the observations and the current value of Θ.

Note that the parameter Θ and the observations $Y_1, \ldots, Y_n$ are conditionally independent, given the non-observed process $X_{[0,T]}$. In particular, from (5), the conditional distribution of Θ given $X_{[0,T]}$, has density, with respect to $\mathcal{L}^d$, proportional to $p_\Theta(\theta)\, g(x_{[0,T]}|\theta)$. The update of the parameter is particularly straightforward when a conjugate prior $p_\Theta(\theta)$ is chosen, so that it is possible to derive analytically the conditional distribution of Θ given $X_{[0,T]}$ and sample directly from it. The second step is computationally more demanding. From (5), the conditional distribution of $X_{[0,T]}$, given parameter and observations, has density, with respect to $\mathbb{W}_{T,\sigma}$, proportional to $g(x_{[0,T]}|\theta)\, l(y_1, \ldots, y_n|x)$, and cannot be sampled directly. An appropriate Metropolis-Hastings step is thus required.

Implementation of the algorithm will necessary involve a discretisation of the diffusion sample path. When the SDE cannot be solved, it is possible to use *Euler-Maruyama approximation*. See for example Chapter 9 in Kloeden and Platen (1992). Alternatively, it may be possible to simulate the diffusion path by means of the exact algorithm described in Beskos, Papaspiliopoulos, Roberts, and Fearnhead (2006), thus avoiding approximation errors.

3.1 Hastings-within-Gibbs algorithm for a latent diffusion model

We now give the details of the Hastings-within-Gibbs algorithm for latent diffusion models.

Just as an example, consider a latent diffusion model with base diffusion which is solution of the SDE

$$dX_t = \theta^\top f(X_t)dt + \sigma dB_t\,, \qquad t \geq 0\,, \qquad X_0 = x_0 \tag{6}$$

with $\theta^\top = (\theta_1, \ldots, \theta_d)$, and $f(x)^\top = (f_1(x), \ldots, f_d(x))$, where $f_i(x)$ is some real-valued function, for $i = 1, \ldots, d$. Let the drift $\theta^\top f(x)$ be such that the regularity conditions mentioned in Section 2 are satisfied. Let the prior for $\Theta = (\Theta_1, \ldots, \Theta_d)$ be multivariate Gaussian, with mean vector and variance matrix

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1d} \\ \lambda_{12} & \lambda_{22} & \cdots & \lambda_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1d} & \lambda_{2d} & \cdots & \lambda_{dd} \end{bmatrix}^{-1}$$

Then, the distribution of Θ, given the diffusion $X_{[0,T]} = x_{[0,T]}$, is still Gaussian, with mean and covariance matrix

$$\mu_x = \Sigma_x \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_d \end{bmatrix} \qquad \Sigma_x = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1d} \\ L_{12} & L_{22} & \cdots & L_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ L_{1d} & L_{2d} & \cdots & L_{dd} \end{bmatrix}^{-1} \tag{7}$$

where, for $i = 1, \ldots, d$ and $j = 1, \ldots, d$,

$$S_i := \frac{1}{\sigma^2} \int_0^T f_i(x_t) dx_t + \sum_{j=1}^{d} \lambda_{ij} \mu_j \qquad L_{ij} := \frac{1}{\sigma^2} \int_0^T f_i(x_t) f_j(x_t) dt + \lambda_{ij}.$$

The update of Θ can thus be performed by sampling directly from this conditional distribution.

The update of the diffusion $X_{[0,T]}$ is less straightforward and requires an appropriate Metropolis-Hastings step. It is possible for example to carry out an independence sampler with proposal distribution given by a Brownian motion starting at x_0. To improve the acceptance rate of the move that update the diffusion path, we apply the following updating strategy. Let $0 = t_1 < \ldots < t_m = T$. Instead of proposing a new diffusion path on the whole interval $[0, T]$, we propose to change the trajectory just on a subinterval $[t_i, t_{i+2}]$, keeping fixed the rest of the diffusion. To ensure continuity of the diffusion path, the proposal distribution, for the new trajectory on the subinterval $[t_i, t_{i+2}]$, is a Brownian bridge $BB_{[t_i,t_{i+2}]}(x_{t_i}, x_{t_{i+2}}) = \{BB_t(x_{t_i}, x_{t_{i+2}}) : t_i \leq t \leq t_{i+2}\}$, having as starting and ending points, respectively, the values $X_{t_i} = x_{t_i}$ and $X_{t_{i+2}} = x_{t_{i+2}}$ of the current diffusion. The proposed diffusion path $x^*_{[0,T]}$ is then given by $\{x^*_t = 1\big(t \notin [t_i, t_{i+2}]\big)x_t + 1\big(t \in [t_i, t_{i+2}]\big)bb_t(x_{t_i}, x_{t_{i+2}}) : t \in [0, T]\}$, where $bb_t(x_{t_i}, x_{t_{i+2}})$ is the realization of the Brownian bridge $BB_{[t_i,t_{i+2}]}(x_{t_i}, x_{t_{i+2}})$. This move is accepted with probability

$$1 \wedge \frac{g\big(bb_{[t_i,t_{i+2}]}(x_{t_i}, x_{t_{i+2}})|\theta\big)}{g(x_{[t_i,t_{i+2}]}|\theta)} \frac{l(y_1, \ldots, y_n|x^*_{[0,y_{[n]}]})}{l(y_1, \ldots, y_n|x_{[0,y_{[n]}]})} \tag{8}$$

where $g(x_{[t_i,t_{i+2}]}|\theta)$ is given by Girsanov's formula restricted to the interval $[t_i, t_{i+2}]$, i.e.

$$g(x_{[t_i,t_{i+2}]}|\theta) = \exp\left\{ \int_{t_i}^{t_{i+2}} \frac{\theta^\top f(X_t)}{\sigma^2} dx_t - \frac{1}{2} \int_{t_i}^{t_{i+2}} \frac{\big(\theta^\top f(X_t)\big)^2}{\sigma^2} dt \right\}.$$

The procedure is iterated for $i = 1, \ldots, m-3$. Note that the different blocks $[t_i, t_{i+2}]$ overlap, so that there are no time instants where the diffusion is kept fixed. For the same reason, the last block $[t_{m-2}, T]$ is updated by means of a Brownian motion $B_{[t_{m-2},T]}(x_{t_{m-2}})$ starting at $X_{t_{m-2}} = x_{t_{m-2}}$, so that the value of the diffusion at T may vary. The acceptance coefficient of the move

that update the last block is the same as in (8), with $[t_i, t_{i+2}] = [t_{m-2}, T]$ and $b_{[t_{m-2},T]}(x_{t_{m-2}})$ in place of $bb_{[t_i,t_{i+2}]}(x_{t_i}, x_{t_{i+2}})$, where $b_{[t_{m-2},T]}(x_{t_{m-2}})$ is the realization of the Brownian motion $B_{[t_{m-2},T]}(x_{t_{m-2}})$.

This idea of updating smaller intervals at a time has been used in Shephard and Pitt (1997) for the simulation of non-Gaussian time series models, and later applied for the simulation of discretely observed diffusions, for example by Elerian, Chib, and Shephard (2001).

In Section 5 and 6 we will carry out simulation studies using this latent diffusion model. Note that the choice of a base diffusion having drift linear in the parameter θ is just due for purposes of exposition. In Section 7 we will indeed analyse a real dataset by a latent diffusion model whose base diffusion has drift which is not linear in θ. Also in that case, the update of the diffusion path will be performed according to the technique described above.

4 Multiple groups of observations

We now discuss a straightforward generalization of the framework developed in the previous sections, and deal with the case of multiple groups of observations, where the observations within each group are taken under homogeneous conditions. Consider for example the case in which different treatments are being administered to different groups of patients in a clinical trial.

Given $\Theta = \theta$, let $X^{[1]}, \dots, X^{[q]}$ be q stochastically independent diffusion processes satisfying (1), and $F_{X^{[1]},h}, \dots, F_{X^{[q]},h}$ the relative random distributions as in (3). Now consider q sequences of observations $(Y_n^{[1]})_n, \dots, (Y_n^{[q]})_n$ such that the random variables in $\big((Y_n^{[1]})_n, \dots, (Y_n^{[q]})_n\big)$ are conditionally independent, given $F_{X^{[1]},h}, \dots, F_{X^{[q]},h}$, and the random variables in $(Y_n^{[k]})_n$ have common distribution $F_{X^{[k]},h}$, for $k = 1, \dots, q$.

The joint distribution of $Y_1^{[1]}, \dots, Y_{n_1}^{[1]}, \dots, Y_1^{[q]}, \dots, Y_{n_q}^{[q]}$, given the diffusions $X_{[0,T_1]}^{[1]} = x_{[0,T_1]}^{[1]}, \dots, X_{[0,T_q]}^{[q]} = x_{[0,T_q]}^{[q]}$, has density, with respect to $\mathcal{L}^n$ (where $n = n_1 + \cdots + n_q$), given by

$$l\big(y_1^{[1]}, \dots, y_{n_1}^{[1]}; \dots; y_1^{[q]}, \dots, y_{n_q}^{[q]} \big| x_{[0,T_1]}^{[1]}, \dots, x_{[0,T_q]}^{[q]}\big) = \prod_{k=1}^{q} l\big(y_1^{[k]}, \dots, y_{n_k}^{[k]} \big| x_{[0,y_{[n_k]}]}^{[k]}\big)$$

where $l\big(y_1^{[k]}, \dots, y_{n_k}^{[k]} \big| x_{[0,y_{[n_k]}]}^{[k]}\big)$ is as in (4). The joint posterior distribution of Θ and $X_{[0,T_1]}^{[1]}, \dots, X_{[0,T_q]}^{[q]}$ has density, with respect to the product measure $\mathcal{L}^d \otimes \mathbb{W}_{T_1,\sigma} \otimes \cdots \otimes \mathbb{W}_{T_q,\sigma}$, given by

$$\begin{aligned} &\pi\big(\theta, x_{[0,T_1]}^{[1]}, \dots, x_{[0,T_q]}^{[q]} \big| y_1^{[1]}, \dots, y_{n_1}^{[1]}; \dots; y_1^{[q]}, \dots, y_{n_q}^{[q]}\big) \\ &= C\, p_\Theta(\theta) \left[\prod_{k=1}^{q} g\big(x_{[0,T_k]}^{[k]} \big| \theta\big)\, l\big(y_1^{[k]}, \dots, y_{n_k}^{[k]} \big| x_{[0,T_k]}^{[k]}\big)\right] \end{aligned} \tag{9}$$

where C is a normalizing constant, and $g\big(x_{[0,T_k]}^{[k]} \big| \theta\big) = \frac{d\mathbb{P}_{T_k,\theta}}{d\mathbb{W}_{T_k,\sigma}}\big(x_{[0,T_k]}^{[k]}\big)$ is given by Girsanov's formula (2).

The contributions of the q groups of observations factorize in (9), and a simple modification of the MCMC algorithm presented in Section 3 may be used to deal with this case. The Hastings-within-Gibbs algorithm for sampling from (9) alternates between

1. simulation of Θ, conditional on the current paths of the q diffusions $X_{[0,T_1]}^{[1]}, \dots, X_{[0,T_q]}^{[q]}$;
2. for each k in $\{1, \dots, q\}$, simulation of $X_{[0,T_k]}^{[k]}$, conditional on the observations $Y_1^{[k]}, \dots, Y_{n_k}^{[k]}$, and the current value of Θ.

Consider, for example, a latent diffusion model with q stochastically independent diffusion processes, $X^{[1]}, \ldots, X^{[q]}$, satisfying the SDE (6). Choose the same multivariate Gaussian prior for Θ that has been used in Section 3.1. Then, the distribution of Θ, given $X^{[1]}_{[0,T_1]} = x^{[1]}_{[0,T_1]}, \ldots, X^{[q]}_{[0,T_q]} = x^{[q]}_{[0,T_q]}$, is still Gaussian, with mean vector and covariance matrix as in (7), but with

$$S_i := \frac{1}{\sigma^2}\left[\sum_{k=1}^{q}\int_0^{T_k} f_i\big(x_t^{[k]}\big)dx_t^{[k]}\right] + \sum_{j=1}^{d}\lambda_{ij}\mu_j \qquad L_{ij} := \frac{1}{\sigma^2}\left[\sum_{k=1}^{q}\int_0^{T_k} f_i\big(x_t^{[k]}\big)f_j\big(x_t^{[k]}\big)dt\right] + \lambda_{ij}$$

for $i = 1, \ldots, d$, $j = 1, \ldots, d$. The update of the parameter Θ can thus be performed by sampling directly from this conditional distribution. The second step may be carried out by q repetitions of the updating mechanism described in Section 3.1.

In Section 7 we will apply this model for multiple groups of observations to analyse a dataset from a clinical trial, that has been considered in a number of papers in the context of survival analysis.

Note that we are here considering a simple hierarchical structure, where inference on the separate groups is linked at the level of the finite dimensional parameter Θ. For some applications this might allow too little borrowing of strength for inference across groups. It would be then of interest to consider a more complex hierarchical structure which allows linking the distributions of the separate groups of observations at an intermediate level. For example, the hazard function of each group could be taken to depend both on a baseline hazard function, and on a group specific hazard function which characterise the idiosyncratic behavior in the group. This would of course call for a more complex MCMC scheme, and care would be needed to insure identifiability of the model.

5 Simulation studies

We show here the implementation of the algorithm described in Section 3, by means of a toy example.

Consider the model based on the diffusion process satisfying the SDE

$$dX_t = \theta_1 \sin(X_t)dt + \theta_2 dt + dB_t\,, \qquad t \geq 0\,, \qquad X_0 = 2 \tag{10}$$

with hazard function $h(u) = u^2$. We simulate observations from this model, for values of the parameters $\theta_1 = -1.4$ and $\theta_2 = -1$, and censoring time $C = 0.9$. In particular, we sample one realization x of the diffusion process satisfying (10), with $\theta_1 = -1.4$ and $\theta_2 = -1$. Then we simulate 200 i.i.d. observations from the corresponding distribution $F_{x,h} = 1 - \exp\left\{-\int_0^t (x_s)^2\, ds\right\}$ and we censor the observations at $C = 0.9$. The diffusion is sampled at intervals of length 0.01, using Euler-Maruyama approximation. Figure 1 shows the corresponding hazards (the squared diffusion) and an histogram of sampled data. The hazard function has a typical shape, first (mainly) increasing and then (mainly) decreasing.

We choose as time horizon of interest $T = 1$. We then run the Hastings-within-Gibbs algorithm under the following specifications. The prior for (θ_1, θ_2) is Gaussian, as in Section 3.1, with $\mu_1 = -1.4$, $\mu_2 = -1$, $\lambda_{11} = \lambda_{22} = 1/5$ and $\lambda_{12} = 0$. The starting values of the parameters are $\theta_1 = \theta_2 = 0$, and the starting diffusion is a Brownian motion, starting at $x_0 = 2$. The diffusion

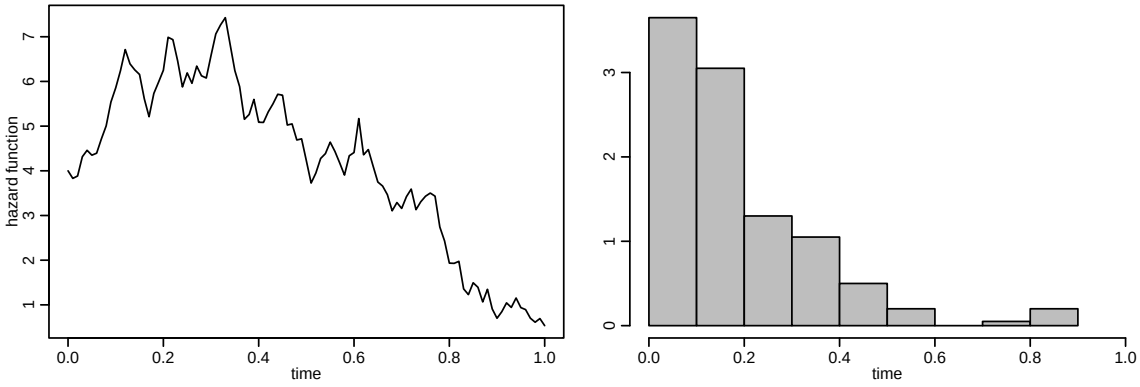

Figure 1: Left: hazard function x^2. Right: histogram of data sampled from F_{x,x^2} with censoring at $C = 0.9$.

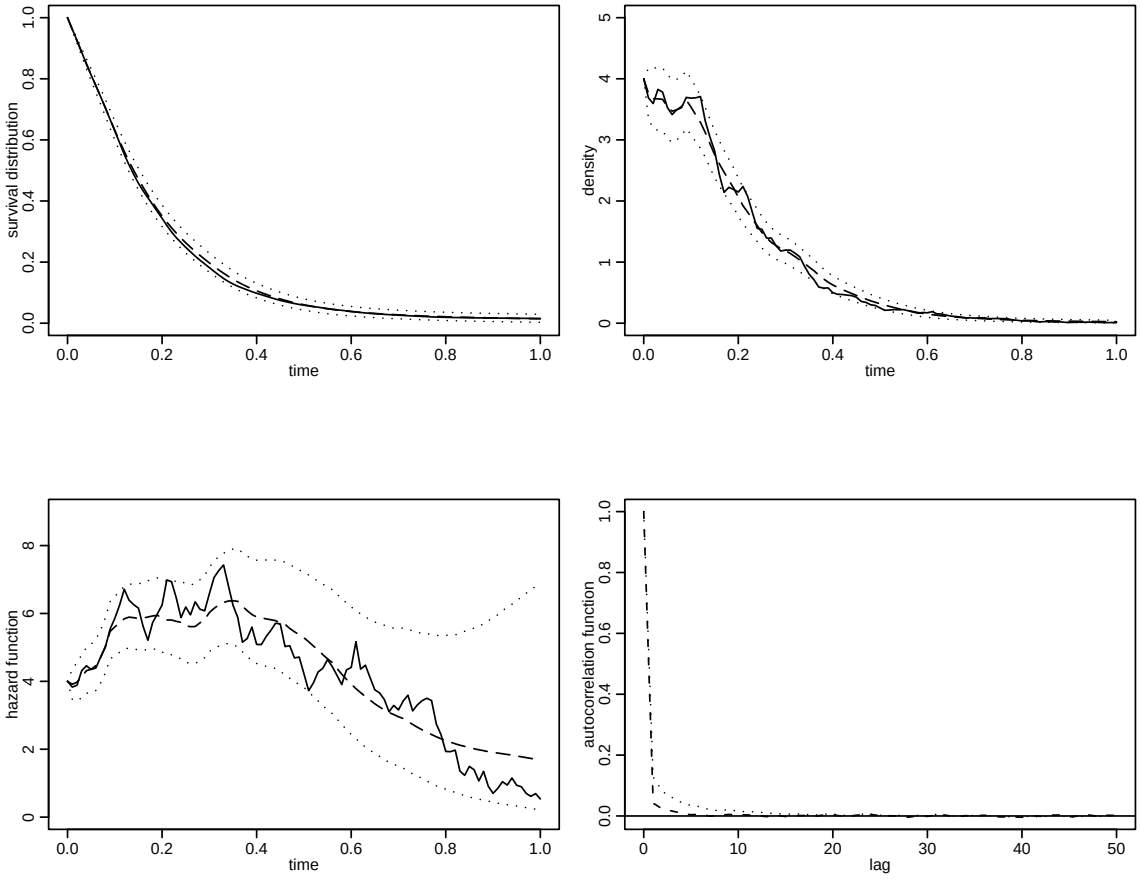

Figure 2: Top left: true survival distribution $1 - F_{x,x^2}$ (solid), together with its posterior mean (dashed) and pointwise approximate 90% highest posterior bands (dotted). Top right: true density (solid), together with its posterior mean (dashed) and pointwise approximate 90% highest posterior bands (dotted). Bottom left: true hazard function x^2 (solid), together with its posterior mean (dashed) and pointwise approximate 90% highest posterior bands (dotted). Bottom right: autocorrelation functions for θ_1 series (dotted) and θ_2 series (dashed).

path is updated on subintervals of length 0.2 at a time. The algorithm is run for 200000 iterations and the first 2000 are discarded as burn in.

Figure 2 shows the estimates of survival distribution, density, and hazard function, based on the MCMC output, together with pointwise approximate 90% highest posterior bands. The true survival distribution and hazard function are also displayed to evidence the good fit of the MCMC estimates. Figure 2 also shows autocorrelation functions for θ_1 and θ_2 series.

6 Partially non-centered reparametrization of model

It may sometime be of interest to consider a finite time horizon T which is significantly bigger than the maximum of the data. In this case the MCMC algorithm described in the previous sections might have poor mixing properties. This problem is evidenced in figure 3. This figure shows the histogram of 200 i.i.d. observations from the distribution $F_{x',h}$, where x' is a new realization of the diffusion process satisfying the same SDE used in Section 5, and also the hazard function h and the censoring time C are the same. This time we choose a longer time horizon $T = 1.8$ (the high number of censored observations, one quarter of the data, suggests that a significant part of the probability mass falls outside the time window where we observed data). We then run the algorithm under the same specifications of Section 5. Figure 3 displays autocorrelation functions for θ_1 and θ_2 series, which are not exponentially decreasing.

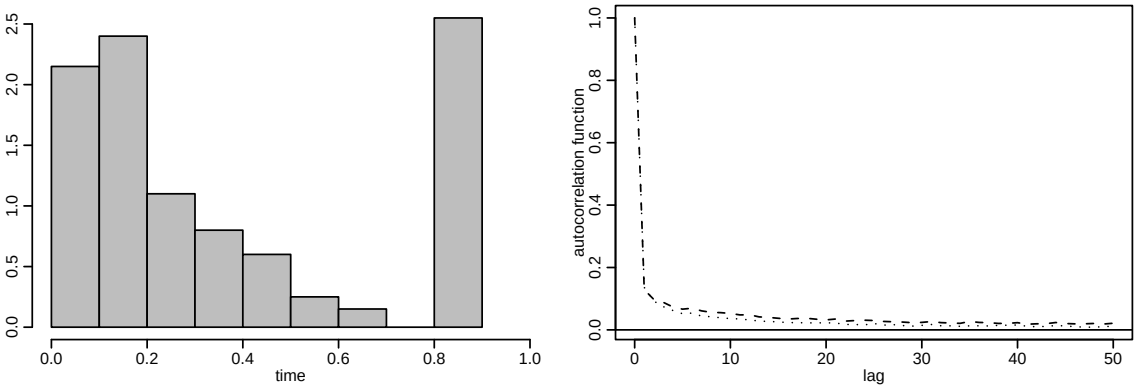

Figure 3: Right: histogram of data sampled from $F_{x',h}$ with censoring at $C = 0.9$. Right: autocorrelation functions for θ_1 and θ_2 series.

To avoid this problem, we propose a modification of the algorithm, based on a reparametrization of the model. Indeed, the performance of MCMC methods, particularly when using Gibbs samplers, depends crucially on the parametrization of the unknown quantities in the hierarchical structure. The issue of reparametrization of the posterior distributions, as to improve convergence properties of the algorithms, has received much attention. See for example Hills and Smith (1992), Gelfand, Sahu, and Carlin (1995), Gelfand, Sahu, and Carlin (1996), and Papaspiliopoulos, Roberts, and Sköld (2003, 2007).

Instead of using the natural parametrization of the model in terms of (Θ, X), the so-called *centered parametrization*, we parametrize it in terms of $(\Theta, \widetilde{X})$, where

$$\widetilde{X}_t = 1(t \leq y_{[n]})\, X_t + 1(t > y_{[n]}) \big[B_t - B_{y_{[n]}}\big] \qquad t \geq 0. \tag{11}$$

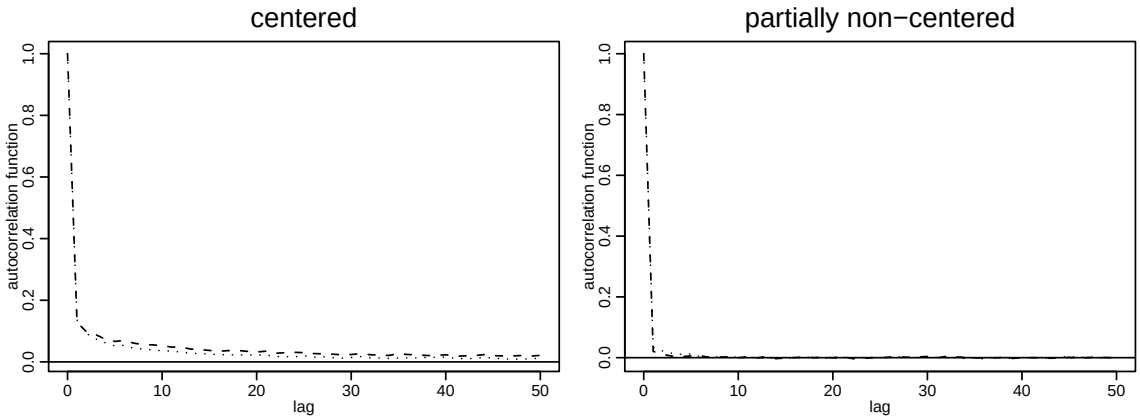

Figure 4: Autocorrelation functions for θ_1 series (dotted) and θ_2 series (dashed), obtained with the algorithm based on the centered parametrization (left) and with the algorithm based on the partially non-centered parametrization (right).

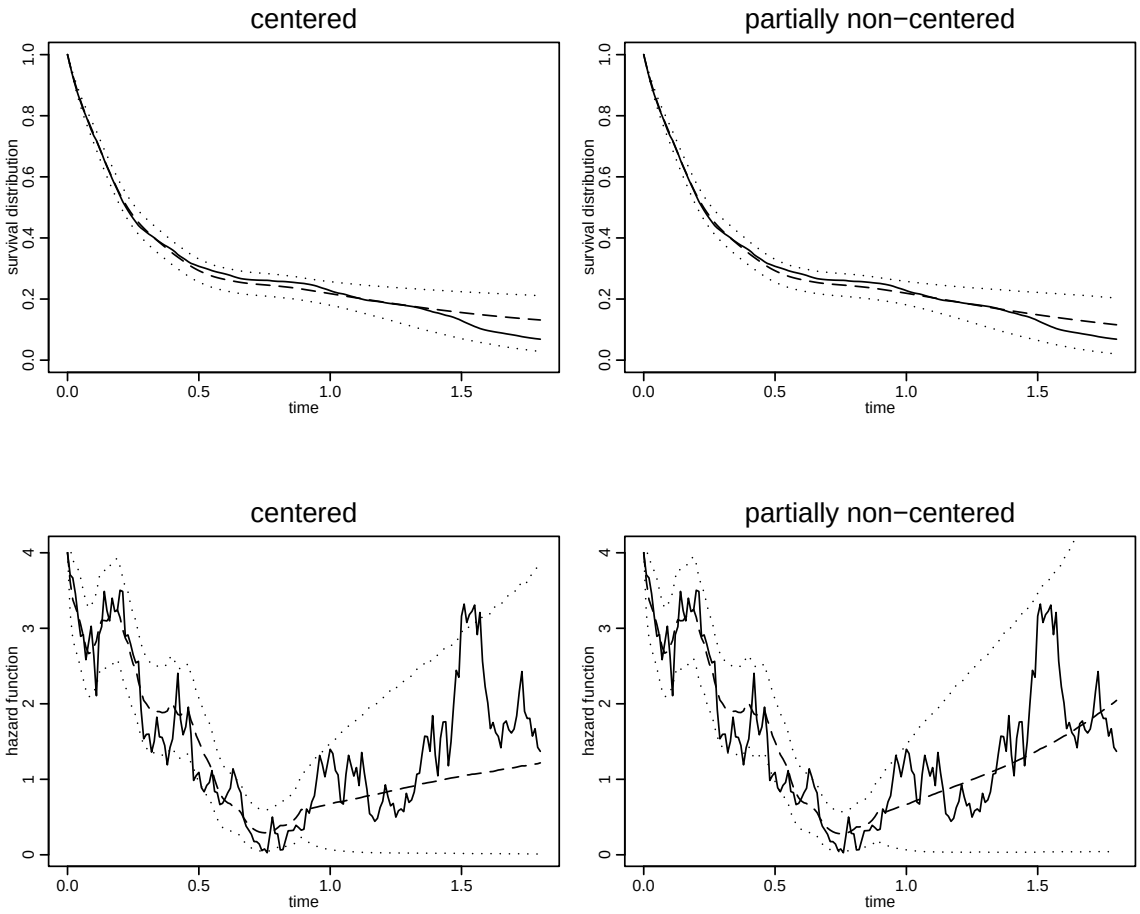

Figure 5: Top: true survival distribution distribution $1 - F_{x',x'2}$ (solid), together with its posterior mean (dashed) and pointwise approximate 90% highest posterior bands (dotted), obtained with the algorithm based on the centered parametrization (left) and with the algorithm based on the partially non-centered parametrization (right). Bottom: true hazard function x'^2 (solid), together with its posterior mean (dashed) and pointwise approximate 90% highest posterior bands (dotted), obtained with the algorithm based on the centered parametrization (left) and with the algorithm based on the partially non-centered parametrization (right).

In the terminology used by Papaspiliopoulos, Roberts, and Sköld (2003), this is called a *partially non-centered parametrization*, the fully *non-centered parametrization* being, in this case, (Θ, B). Using the parametrization (11) expresses the idea that the data carry no information on the diffusion after the maximum data point. The diffusion X can then be reconstructed as function of Θ, $\widetilde{X}$ and $y_1, \ldots, y_n$, by

$$\begin{cases} X_t = \widetilde{X}_t & 0 \le t \le y_{[n]} \\ dX_t = \beta(X_t, \Theta)dt + \sigma d\widetilde{X}_t & t \ge y_{[n]}. \end{cases}$$

The joint posterior distribution of Θ and $\widetilde{X}$ has density, with respect to the product measure $\mathcal{L}^d \otimes \mathbb{W}_\sigma$, given by

$$\pi\big(\theta, \tilde{x} \big| y_1, \ldots, y_n\big) = C\, p_\Theta\big(\theta\big)\; g\big(x_{[0,y_{[n]}]} \big| \theta\big)\; l\big(y_1, \ldots, y_n \big| \theta, x_{[0,y_{[n]}]}\big) \tag{12}$$

where C is a normalizing constant, and $g\big(x_{[0,y_{[n]}]}\big|\theta\big) = \frac{d\mathbb{P}_{y_{[n]},\theta}}{d\mathbb{W}_{y_{[n]},\sigma}}(x_{[0,y_{[n]}]})$. Note that (12) characterizes the posterior distribution of $\widetilde{X}$, and thus the posterior distribution of the diffusion X, over the whole positive half-line.

It is possible to simulate from (12) by means of a Gibbs sampler quite similar to the one described in Section 3.1. In the first step we simulate Θ conditionally on $\widetilde{X}_{[0,y_{[n]}]} \equiv X_{[0,y_{[n]}]}$. In the second step, we simulate $\widetilde{X}$ over the time interval of interest $[0, T]$, conditionally on Θ. In this case we use a proposal distribution which is a Brownian motion starting at x_0, over the time interval $[0, y_{[n]}]$, and a Brownian motion starting at 0, over the time interval $[y_{[n]}, T]$. On $[0, y_{[n]}]$ we follow again the updating strategy, with the overlapping Brownian bridges, described in Section 3.1. When reconstructing the diffusion $X_{[0,T]}$, from Θ and $\widetilde{X}_{[0,T]}$, we are careful to preserve continuity of the diffusion path at time $y_{[n]}$. Details are omitted.

The algorithm based on the reparametrization (11) is completely robust to the choice of T, since the update of the parameter Θ, conditionally on $\widetilde{X}$, only involve $\widetilde{X}_{[0,y_{[n]}]}$. Moreover, if the algorithm has been run with a certain choice of T, and it later becomes of interest a longer time horizon T', with $T' > T$, we can obtain sample paths of $\widetilde{X}_{[T,T']}$ by additional post hoc simulation, using the values of Θ and X_T that have been sampled along the chain. With the centered parametrization it is instead necessary to run again the algorithm from the beginning, changing the time window from $[0, T]$ to $[0, T']$.

Figures 4 and 5 compares mixing and MCMC estimates obtained with the algorithms based on the centered parametrization and on the partially non-centered parametrization, for the data set corresponding to figure 3. The specifications of the two algorithms are as in Section 5. Note that the hazard function is bathtub shaped. Hazard functions with such shapes are quite common in survival analysis (think, for instance, to human mortality).

7 Application to real data

In this section we apply our latent diffusion model for multiple groups of observations to a dataset from a clinical trial, that has been considered in a number of papers in the context of survival analysis, among which Gehan (1965), Cox (1972), Wei (1984) and Xu and O'Quigley (2000) in the non-Bayesian literature, and Kalbfleisch (1978), Laud, Damien, and Smith (1998) and Damien and

Walker (2002) in the Bayesian one. In the trial, reported by Freireich (1963), 6-mercaptopurine (6-MP) was compared to a placebo in the maintenance of remission in acute leukemia. The following lengths of remission in weeks were recorded for 42 patients, half of which treated with the 6-MP drug and half with the placebo (a + sign indicates a censored observation):

6-MP: $6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+$

placebo: $1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.$

We consider the model for multiple groups of observations (here 2 groups, 6-MP drug and placebo), based on the diffusion process satisfying the SDE

$$dX_t = \theta_1 \left(sign(X_t)\right) |X_t|^{\theta_2} \, dt \; + \; \sigma \, dB_t \,, \qquad t \geq 0 \,, \qquad X_0 = x_0 \tag{13}$$

where

$$sign(u) = 1 \begin{cases} 1 & \text{if } u > 0 \\ -1 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \end{cases}$$

with hazard function $h(u) = |u|$. Note that when $\sigma = 0$, this is equivalent to the Weibull model. Indeed, for $\sigma = 0$, the SDE (13) reduces to the differential equation $dX_t = \theta_1 X_t^{\theta_2} dt$, which has solution $X_t = \left(\theta_1(1-\theta_2)\right)^{\frac{1}{1-\theta_2}} \, t^{\frac{1}{1-\theta_2}}$, so that the hazard $h(X_t)$ is proportional to a power of t, as the hazard of Weibull distribution. Thus, the model based on the diffusion process (13), with hazard function $h(u) = |u|$, is a stochastic perturbation around a central Weibull model.

We express the data as fractions of one year, and choose as time horizons of interest $T_1 = T_2 = 0.75$, corresponding to 9 months (39 weeks). We take Θ_1 and Θ_2 a priori independent, with a Gaussian prior distribution for Θ_1, with mean $\mu = 0$ and variance $1/\lambda = 5$, and a uniform prior over $[0, 1]$ for Θ_2. We moreover set $x_0 = 0.8$ and $\sigma = 8$. We then run the Hastings-within-Gibbs algorithm based on the partially non centered parametrization. The update of Θ_1 is performed by sampling directly from the conditional distribution Θ_1 given $\Theta_2, \widetilde{X}^{[1]}, \widetilde{X}^{[2]}$, which is still Gaussian with mean $\frac{S+\lambda\mu}{L+\lambda}$ and variance $\frac{1}{L+\lambda}$, where

$$S := \frac{1}{\sigma^2} \left[\sum_{j=1}^{2} \int_0^{y_{[n_j]}} \left((sign(x_t^{[j]})) \, |x_t^{[j]}|^{\theta_2} \right) dx_t^{[j]} \right] \qquad L := \frac{1}{\sigma^2} \left[\sum_{j=1}^{2} \int_0^{y_{[n_j]}} \left(|x_t^{[j]}|^{\theta_2} \right)^2 dt \right]$$

For the update of Θ_2 we use an independence sampler with a Beta proposal distribution, with parameters $(1/2, 1/2)$. The update of $\widetilde{X}^{[1]}$ and $\widetilde{X}^{[2]}$ is carried out as described in the previous sections. The algorithm is run for 200000 iterations and the first 2000 are discarded as burn in.

Figure 6 displays the MCMC estimates of the survival distributions of the two groups, 6-MP drug and placebo, together with the relative Kaplan-Meier curves. Note that the MCMC estimates of the two survival distributions are closer one another than the two Kaplan-Meier curves, thus showing borrowing of strength for inference among the two groups. Hence, the latent diffusion model, that gains much flexibility over a fully parametric model by introducing randomness around it, does not suffer from the opposite problem of being too data-driven. Figure 6 also displays the MCMC estimates of the hazards of the two groups.

We could now verify the efficacy of 6-MP drug treatment as proposed in Damien and Walker (2002). In particular, under the hypothesis that 6-MP drug is inefficient, we would regard all patients as belonging to 1 single group, instead of 2. We could then implement the latent diffusion

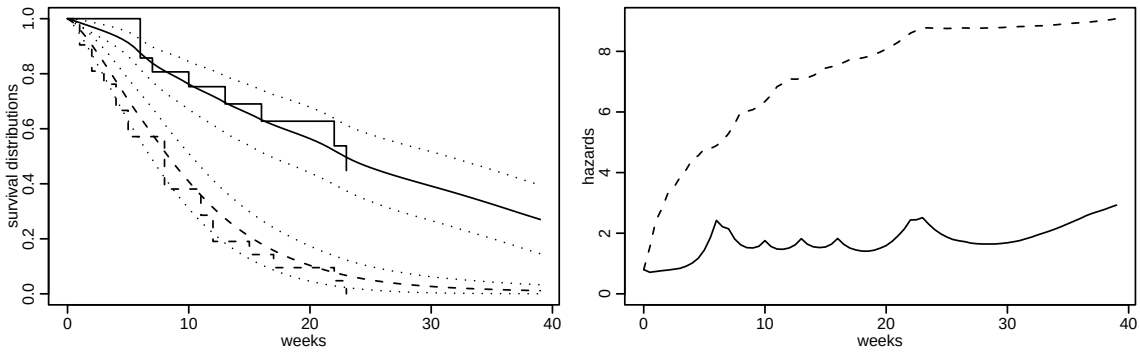

Figure 6: Left: posterior mean survival distributions and pointwise approximate 90% highest posterior bands, for the group of patients treated with 6-MP drug (solid) and for the group of patients treated with the placebo (dashed), together with corresponding Kaplan-Meier curves. Right: posterior mean hazards for the group of patients treated with 6-MP drug (solid) and for the group of patients treated with the placebo (dashed).

model based on (13), but with just 1 diffusion process. Call Model_1 the model where all patients belong to 1 single group (corresponding to the hypothesis H_1 of null efficacy of 6-MP drug), and call Model_2 the one considered above (corresponding to the hypothesis H_2 of efficacy of 6-MP drug). If the a priori probabilities of hypothesis H_1 and H_2 are set equal to 0.5, the Bayes Factor

$$\text{BF} = \frac{\text{probability density of data under model } \text{M}_1}{\text{probability density of data under model } \text{M}_2}$$

gives the posterior odds in favor of H_1. As expected, the computed Bayes Factor (BF$= 9 \times 10^{-6}$) gives a strong evidence of the efficacy of 6-MP drug.

8 Generalization to the case of unknown diffusion coefficient

An important generalization of the model we have considered so far consist in considering diffusion processes with unknown diffusion coefficient σ, since σ describes a natural measure of prior uncertainty. We briefly discuss how to deal with this case.

Let Σ be a real random variable. Given $\Theta = \theta$ and $\Sigma = \sigma$, consider the scalar diffusion process X solution of the SDE (1), and denote by $\mathbb{P}_{T,\theta,\sigma}$ the law of $X_{[0,T]}$. Let $p_\Sigma(\cdot)$ be the prior density, with respect to $\mathcal{L}$, of Σ (for simplicity, we take Θ and Σ be stochastically independent a priori). Then, the joint posterior distribution of $(\Theta, \Sigma, X_{[0,T]})$ has density, with respect to $\mathcal{L}^{d+1} \otimes \mathbb{W}_{T,\sigma}$, given by

$$\pi(\theta, \sigma, x_{[0,T]}|y_1, \ldots, y_n) = C\, p_\Theta(\theta)\, p_\Sigma(\sigma)\, g(x_{[0,T]}|\theta, \sigma)\, l(y_1, \ldots, y_n|x_{[0,y_{[n]}]}) \tag{14}$$

where C is a normalizing constant, and $g(x_{[0,T]}|\theta, \sigma) := \frac{d\mathbb{P}_{T,\theta,\sigma}}{d\mathbb{W}_{T,\sigma}}(x_{[0,T]})$ is given by Girsanov's formula (2).

The quadratic variation of a diffusion processes, having diffusion coefficient σ, satisfies

$$\lim_{m\to\infty} \sum_{i=1}^{m} (X_{ti/m} - X_{t(i-1)/m})^2 = t\sigma^2 \qquad \mathbb{W}_{T,\sigma} - \text{a.s. for all } t.$$

Therefore, the conditional distribution of Σ, given the diffusion $X_{[0,T]}$, degenerates to a point mass, and Σ is completely determined by the diffusion path. In practice, we cannot simulate the diffusion

path in continuous time, but just at discrete time instants. Anyway, the finer the time discrete approximation $\{X_{iT/m} : i = 1, \ldots, m\}$ of the diffusion $X_{[0,T]}$, the stronger becomes the dependence between $\{X_{iT/m} : i = 1, \ldots, m\}$ and Σ. Consider the algorithm for the simulation from (14), that alternates between

1. simulation of Θ, conditional on the current value of Σ and the current path of $X_{[0,T]}$;
2. simulation of Σ, conditional on the current value of Θ and the current path of $X_{[0,T]}$;
3. simulation of $X_{[0,T]}$, conditional on the observations and the current values of Θ and Σ.

The finer the approximation of the diffusion path, the worse the convergence of the algorithm becomes. In the limiting case $m = \infty$ (that is, if the diffusion process could be simulated in continuous time), this scheme would be reducible. See Roberts and Stramer (2001). An alternative way to see this problem is to note that the collection of measures $\{\mathbb{W}_{T,\sigma} : \sigma \in \mathbb{R}\}$ are mutually singular, and therefore so are the measures $\{\mathbb{P}_{T,\theta,\sigma} : \sigma \in \mathbb{R}\}$.

In this case, the need for a different parametrization of the model is thus compelling. Following Roberts and Stramer (2001), we parametrize the model in terms of $(\Theta, \Sigma, \dot{X})$, where $\dot{X}_t = (X_t - X_0)/\Sigma$. By *Îto's formula*,

$$d\dot{X}_t = \frac{\beta(\dot{X}_t, \Theta)}{\Sigma} dt + dB_t\,, \qquad t \geq 0\,, \qquad \dot{X}_0 = 0.$$

The distribution of $\dot{X}_{[0,T]}$ depends on Σ, but any realization of $\dot{X}_{[0,T]}$ contains only finite information about Σ. Another possible reparametrization of the model, along the lines of Section 6, could be in terms of $(\Theta, \Sigma, \ddot{X})$, with

$$\ddot{X}_t = 1(t \leq y_{[n]})\, \dot{X}_t + 1(t > y_{[n]}) \big[B_t - B_{y_{[n]}}\big] \qquad t \geq 0.$$

MCMC algorithms based on these reparametrizations can be obtained as simple modifications of the ones previously described.

Consider the toy example described in Section 5, and assume the same model, but let the diffusion process have an unknown diffusion coefficient. Let the prior for this coefficient be exponential with mean 1. Figure 7 displays the results obtained with the MCMC algorithm based on the reparametrization $(\Theta, \Sigma, \dot{X})$. Specification of the algorithm are as in Section 5. Note that the mixing for σ is slow relatively to the very good mixing for θ_1 and θ_2, but this does not prevent good estimates of the survival distribution, density and hazard being obtained. Slow mixing for σ could be probably improved by a further reparametrization of the model.

Alternatively to the case of unknown diffusion coefficient, it would be possible to consider models based on diffusion processes having $\sigma = 1$, but with hazard function $h(\Gamma, X)$, where Γ is a random parameter. Also in this case, a reparametrization of the model would be necessary.

9 Discussion

In this paper we have described a latent diffusion model for event history analysis, considering both the cases of a single group of observations and of multiple groups of observations. We have shown that the model can be efficiently treated by means of MCMC techniques. All analyses presented are computationally feasible within R©.

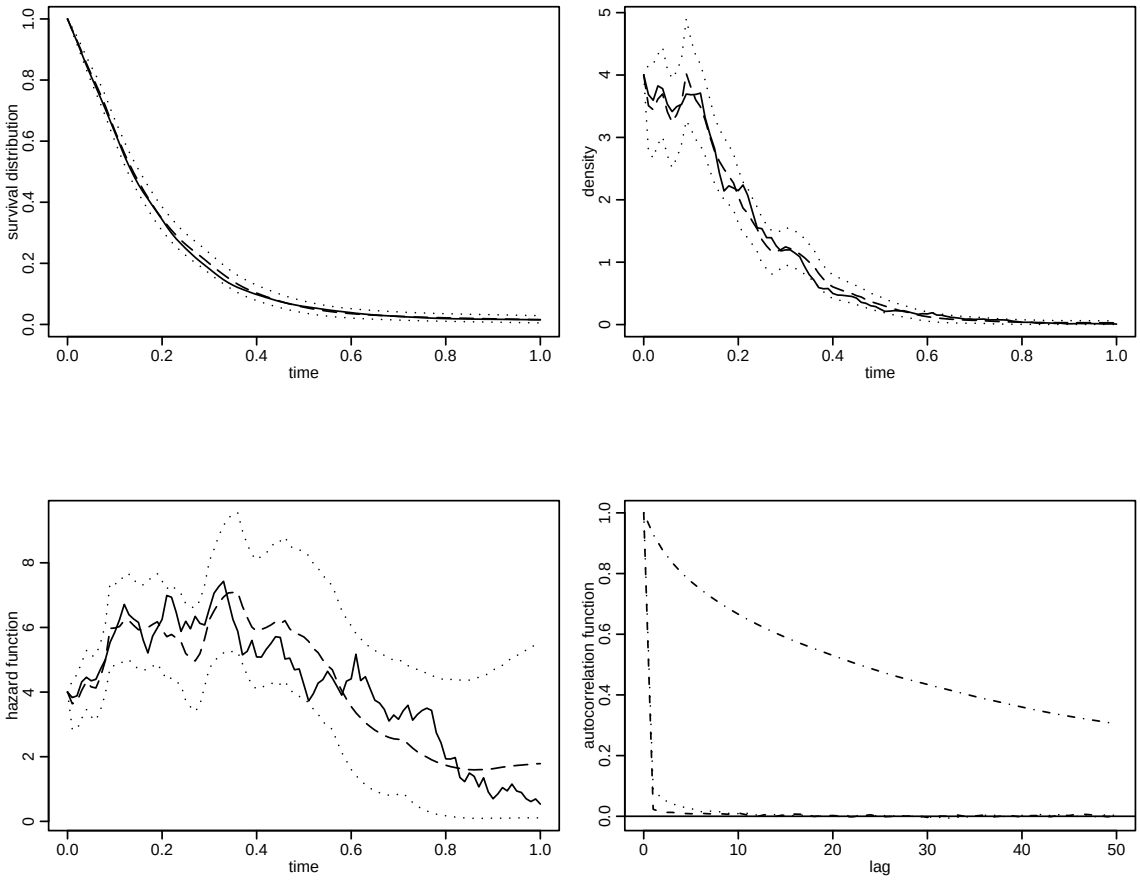

Figure 7: As in Figure (2), but for the model with unknown diffusion coeffient. Bottom right plot also displays autocorrelation function for σ series (dotdash line).

Covariates can be included in this framework in a very natural way, as influencing directly the underlying diffusion. If Z_t is the covariate process, we could for example consider a model based on the diffusion satisfying the SDE

$$dX_t = \beta(X_t, Z_t, \theta) + \sigma dB_t \qquad t \geq 0$$
$$X_0 = x_0(Z_0).$$

In particular, we could follow what Aalen and Gjessing (2001) suggest for barrier hitting models. Namely, those covariates which represent measures of how far the underlying process, that leads to the event, has advanced (such as staging measures in cancer) may be taken to influence the starting point of the diffusion; those covariates which instead represent causal influence on the development of the process may be taken to influence the drift of the diffusion. See Aalen and Gjessing (2001) for interesting discussions about this choice.

An interesting generalization of the model would be to consider random probabilities based on jump diffusion processes. As noticed in Section 2, the cumulative hazard functions, associated with random probabilities based on diffusions, are smooth, being the integrals of continuous processes. By replacing the diffusion process with a jump diffusion process it would be possible to capture sudden changes in the behavior of cumulative hazards, that might be due to some kind of shock

experienced by the population.

Acknowledgments

We would like to thank Robin Henderson and Piercesare Secchi for useful comments, and Omiros Papaspiliopoulos and Alexandros Beskos for their help with the programming. The second author acknowledge funding by EC Marie Curie Training Site Human Potential Programme, to visit the Department of Mathematics and Statistics, Lancaster University, and by the Centre for Research in Statistical Methodology (CRiSM), Warwick University.

References

Aalen, O. O. and Gjessing, H. K. (2001), "Understanding the shape of the hazard rate: a process point of view," *Statist. Sci.*, 16, 1–22, with comments and a rejoinder by the authors.

— (2004), "Survival models based on the Ornstein-Uhlenbeck process," *Lifetime Data Anal.*, 10, 407–423.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006), "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes," *J. R. Statist. Soc. B*, to appear.

Cox, D. R. (1972), "Regression models and life-tables," *J. Roy. Statist. Soc. Ser. B*, 34, 187–220, with discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

Damien, P. and Walker, S. (2002), "A Bayesian non-parametric comparison of two treatments," *Scand. J. Statist.*, 29, 51–56.

Doksum, K. (1974), "Tailfree and neutral random probabilities and their posterior distributions," *Ann. Probability*, 2, 183–201.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood inference for discretely observed nonlinear diffusions," *Econometrica*, 69, 959–993.

Freireich, E. O. (1963), "The effect of 6 mercaptopurine on the duration of steroid induced remission in acute leukemia," *Blood*, 21, 699–716.

Gehan, E. A. (1965), "A generalized Wilcoxon test for comparing arbitrarily singly-censored samples," *Biometrika*, 52, 203–223.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient parameterisations for normal linear mixed models," *Biometrika*, 82, 479–488.

— (1996), "Efficient parametrizations for generalized linear mixed models," in *Bayesian statistics, 5 (Alicante, 1994)*, New York: Oxford Univ. Press, Oxford Sci. Publ., pp. 165–180.

Hills, S. E. and Smith, A. F. M. (1992), "Parameterization issues in Bayesian inference," in *Bayesian statistics, 4 (Peñíscola, 1991)*, New York: Oxford Univ. Press, pp. 227–246.

Hjort, N. L. (1990), "Nonparametric Bayes estimators based on beta processes in models for life history data," *Ann. Statist.*, 18, 1259–1294.

Kalbfleisch, J. D. (1978), "Non-parametric Bayesian analysis of survival time data," *J. Roy. Statist. Soc. Ser. B*, 40, 214–221.

Kloeden, P. E. and Platen, E. (1992), *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*, Berlin: Springer-Verlag.

Laud, P. W., Damien, P., and Smith, A. F. M. (1998), "Bayesian nonparametric and covariate analysis of failure time data," in *Practical nonparametric and semiparametric Bayesian statistics*, New York: Springer, vol. 133 of *Lecture Notes in Statist.*, pp. 213–225.

Myers, L. E. (1981), "Survival functions induced by stochastic covariate processes," *J. Appl. Probab.*, 18, 523–529.

Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2003), "Non-centered parameterizations for hierarchical models and data augmentation," in *Bayesian statistics, 7 (Tenerife, 2002)*, New York: Oxford Univ. Press, pp. 307–326, with a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors.

— (2007), "A general framework for the parametrization of Hierarchical models," *Statit. Sci.*, 22, 59–73.

Roberts, G. O. and Stramer, O. (2001), "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm," *Biometrika*, 88, 603–621.

Rogers, L. C. G. and Williams, D. (2000), *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge: Cambridge University Press, itô calculus, Reprint of the second (1994) edition.

Shephard, N. and Pitt, M. K. (1997), "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 84, 653–667.

Wei, L. J. (1984), "Testing goodness of fit for proportional hazards model with censored observations," *J. Amer. Statist. Assoc.*, 79, 649–652.

Woodbury, M. A. and Manton, K. G. (1977), "A random-walk model of human mortality and aging," *Theoret. Population Biology*, 11, 37–48.

Xu, R. and O'Quigley, J. (2000), "Proportional hazards estimate of the conditional survival function," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62, 667–680.

Yashin, A. I. (1985), "Dynamics of survival analysis: conditional Gaussian property versus the Cameron-Martin formula," in *Statistics and control of stochastic processes (Moscow, 1984)*, New York: Optimization Software, Transl. Ser. Math. Engrg., pp. 466–485.

Yashin, A. I. and Vaupel, J. W. (1986), "Measurement and estimation in heterogeneous populations," in *Immunology and epidemiology (Mogilany, 1985)*, Berlin: Springer, vol. 65 of *Lecture Notes in Biomath.*, pp. 198–206.