A latent process model for joint modeling of events and marker.

Reza Hashemi, Hélène Jacqmin-Gadda, Daniel Commenges

A Latent Process Model for Joint Modeling of Events and Marker

R. Hashemi, H. Jacqmin-Gadda and D. Commenges

INSERM E03 38, 146 rue Léo Saignat, 33076 Bordeaux, France
e-mail: daniel.commenges@isped.u-bordeaux2.fr

Abstract

The paper formulates joint modeling of a counting process and a sequence of longitudinal measurements, governed by a common latent stochastic process. The latent process is modeled as a function of explanatory variables and a Brownian motion process. The conditional likelihood given values of the latent process at the measurement times, has been drawn using Brownian bridge properties; then integrating over all possible values of the latent process at the measurement times leads to the desired joint likelihood. An estimation procedure using joint likelihood and a numerical optimization is described. The method is applied to the study of cognitive decline and Alzheimer's disease.

Key words: latent process, Brownian motion, joint model, dementia, Alzheimer's disease

1 Introduction

A detailed description of the evolution of the health of subjects involves considering both continuous processes and events. This is also true when considering an industrial item for which a continuous degradation process precedes the failure. One example in epidemiology is the evolution of the number of CD4 lymphocytes and the onset of AIDS, or of death, in HIV infected patients (see Wulfsohn and Tsiatis, 1997; Faucett and Thomas, 1996). Another example, which will serve as an illustration of our model, is the evolution of cognitive abilities and onset of dementia. The aims of joint models are in particular a better description of the link of the two kinds of observations, the increase in power to detect factors influencing the evolution, the prediction of the event using the observation of the marker.

An approach of joint modeling tries to directly relate the distribution of the marker and of the time to event (Hogan and Laird, 1997; Faucett and Thomas, 1996). It however seems more interesting to tackle the problem in a stochastic processes framework in which a better description of the dynamic of the phenomena is possible. Stochastic processes were introduced in the modeling of quantitative longitudinal data by Diggle (1988). A joint model based on latent stochastic processes was proposed by Henderson, Diggle and Dobson (2000). In this model the latent process acts as a time dependent variable in a proportional hazard model for the event. A different approach defines the event as the crossing of a barrier by a latent process (Cox, 1999). Aalen and Gjessing (2001) studied the shape of hazard functions for events defined as the hitting time of a barrier by a Wiener process (equivalently a Brownian motion), possibly with drift and random effects. Whitmore, Crowder and Lawless (1998) proposed a joint model in which the event was defined as the crossing of a barrier by a latent Wiener process; the model is

quite flexible because the marker is determined by another process which may be correlated to the first. Lee, DeGruttola and Schoenfeld (2000) extended this model to take covariates into account. However only the case where one observation of the marker, at the time of censoring, was available has been dealt with.

The aim of this paper is to pursue the latter approach of modeling, for mainly two reasons: defining an event in this way fits conceptually well to the modeling of dementia (dementia can really be considered as defined as having cognitive abilities below a certain level); mathematically the model is nice because the hitting probabilities can be computed by relatively simple formulas for the Brownian motion. Nevertheless, considerable numerical difficulties arise, but they can be dealt with. We shall present a joint model for a marker and an event based on a latent process modeled linearly as a function of explanatory variables having fixed or random effects and driven by a Brownian motion. The event of interest occurs when the latent process hits a barrier. We may have either exact or interval-censored observations of the hitting time and the process is indirectly observed at discrete times by a marker.

This model is described in section 2. In section 3 the likelihood for the observations is given; the case where the observations are left-truncated is treated. In section 4 an algorithm, based on a version of the Newton-Raphson algorithm, is proposed. A simulation study is presented in section 5, which in particular illustrates the gain in precision obtained by using the joint information as compared to using only the marker information. In section 6 a joint model of the onset of dementia and observation of a psychometric test is fitted on the data of the Paquid cohort study. Section 7 concludes.

2 Model

2.1 Definition of the model

We consider that all observations of markers and events depend on latent processes $\Lambda_i(t)$, $i = 1, \cdots, n$. In our application we think of the latent process as the cognitive ability of the subject; it is modeled as a function of explanatory variables and a stochastic term, $W_i(t)$, which is taken to be a standard Brownian motion (a variance parameter would not be identifiable). The model is as follows:

$$\Lambda_i(t) = (\mathbf{Z}_i^1\, \alpha + \mathbf{Z}_i^2\, \mathbf{a}_i) + (\mathbf{Z}_i^3\, \beta + \mathbf{Z}_i^4\, \mathbf{b}_i)\, t + W_i(t) \tag{1}$$

where α and β are vectors of unknown regression parameters associated with explanatory variables $\mathbf{Z}_i^1$ and $\mathbf{Z}_i^3$, $\mathbf{a}_i$ and $\mathbf{b}_i$ are vectors of random effects associated with explanatory variables $\mathbf{Z}_i^2$ and $\mathbf{Z}_i^4$ respectively. The random effects $\mathbf{a}_i$ and $\mathbf{b}_i$ are assumed to be independent of $W_i(t)$, $i = 1, \cdots, n$, and normally distributed with zero means and variance

$$\Gamma = \begin{pmatrix} \Gamma_{\mathbf{a}} & \Gamma_{\mathbf{ab}} \\ \Gamma_{\mathbf{ba}} & \Gamma_{\mathbf{b}} \end{pmatrix}.$$

The relation between the latent process and its indirect measurement at time t_{ij} is modeled as:

$$X_{ij} = \gamma_0 + \gamma_1\, \Lambda_i(t_{ij}) + \varepsilon_{ij} \qquad ; j = 1, \cdots, r_i$$

where γ_0 and γ_1 are parameters and ε_{ij} are measurement errors which are assumed to be identically and independently normally distributed with zero mean and variance σ_e^2. The process of observation times is independent of $\Lambda_i(t)$.

A subject becomes ill when the latent process reaches the threshold η for the first time, and we denote this time T_η^i; more formally $T_\eta^i = \inf\{t \geq 0; \Lambda_i(t) \leq \eta\}$ is the first passage time to η. The clinical state of the subjects is described by $D_i(t)$ with $D_i(t) = 0$ if subject is healthy, $D_i(t) = 1$ if ill, so $D_i(t) = I_{\{T_\eta^i \leq t\}}$. D_i may be observed either in continuous time or in discrete time.

Consider first the often more realistic case of observation at discrete times t_{ij}, $j = 1, \cdots, r_i$ already defined. In that case we do not observe exactly the first passage time T_η^i but we observe whether $T_\eta^i \leq t_{ij}, j = 1, \cdots, r_i$. There are two possible kinds of observations: subjects who are observed ill at t_{im_i} (thus $t_{im_i} = \min\{t_{ij} : D_i(t_{ij}) = 1,\ j = 1, \cdots r_i\}$), and subjects who remain healthy until t_{ir_i}. It is easy to modify formulas, if there are non-informative missing values, either in X_i, or in D_i. We will also have in the application an artificial missing data mechanism by ignoring the observation of X_i after t_{im_i}. This will make the linear trend assumption more tenable.

In the case where D_i is observed in continuous time, and the event has occured before t_{ir_i}, T_η^i is observed.

A particular feature of the model with random intercept is that there is a non-null probability of being ill at time zero. However this fits with our application on dementia because we take the origin of time at 58 years, an age at which the probability of being demented is non-null.

We consider the case where the sample may be selected by a truncation mechanism; subjects can be selected in the sample only if $D_i(t_{i1}) = 0$, that is, diseased subjects at t_{i1} are excluded.

2.2 Hazard rate

In our model marginally relative to the random effects, the hazard rate for a subject with explanatory variables $\mathbf{Z}_i^1, \cdots, \mathbf{Z}_i^4$, can be written as the ratio of the marginal density over the marginal survival, conditional on these quantities being different from zero:

$$h_i(t) = \frac{\int f(t, \mu_i(\mathbf{b}_i), \psi_i(\mathbf{a}_i), \eta) dP_{\mathbf{a}_i, \mathbf{b}_i}}{\int [1 - F(t, \mu_i(\mathbf{b}_i), \psi_i(\mathbf{a}_i), \eta)] dP_{\mathbf{a}_i, \mathbf{b}_i}}, \quad t > 0, \tag{2}$$

where $f(.)$ is the density of hitting time of the Brownian motion with drift μ to threshold η, (see, e.g. Karatzas and Shreve, 1991) also known as the p.d.f. of the inverse-Gaussian distribution:

$$f(t, \mu, \psi, \eta) = \frac{|\eta - \psi|}{\sqrt{2\,\pi\,t^3}} \exp\left[-\frac{(\eta - \psi - \mu t)^2}{2t}\right]; \quad t > 0, \eta < \psi, \mu \in \mathbb{R}, \tag{3}$$

and,

$$F(t, \mu, \psi, \eta) = 1 - \Phi\left(\mu\sqrt{t} + \frac{\psi - \eta}{\sqrt{t}}\right) + e^{-2\mu(\psi-\eta)}\Phi\left(\mu\sqrt{t} - \frac{\psi - \eta}{\sqrt{t}}\right), \tag{4}$$

where Φ is the cumulative standard normal distribution. $F(t, \mu, \psi, \eta)$ is the probability that a Brownian motion with drift μ starting at $\psi > \eta$ reaches the level η before t: if $\mu \leq 0$, $F(\infty, \mu, \psi, \eta) = 1$; if $\mu > 0$ $F(\infty, \mu, \psi, \eta) = e^{-2\mu(\psi-\eta)}$. For subject i the starting point is $\psi_i(\mathbf{a}_i) = \mathbf{Z}_i^1\,\alpha + \mathbf{Z}_i^2\,\mathbf{a}_i$ and the drift is $\mu_i(\mathbf{b}_i) = \mathbf{Z}_i^3\,\beta + \mathbf{Z}_i^4\,\mathbf{b}_i$. Marginal density and survival for $t > 0$ are obtained by integration of their conditional counterparts given the random effects, relatively to the distribution of the random effects conditional on $\psi_i(\mathbf{a}_i) > \eta$. This integration can be done numerically, using for instance a mean of the integrand term for values of $\mathbf{a}_i, \mathbf{b}_i$ generated from their multivariate normal distribution and rejecting the draws such that $\psi_i(\mathbf{a}_i) \leq \eta$.

3 Likelihood

3.1 Likelihood ignoring the selection of the sample

Consider first the case where D_i is observed in discrete time. Using the fact that $\mathbf{X}_i = (X_{i1}, \cdots, X_{ir_i})^T$ and $\mathbf{D}_i = (D_{i1}, \cdots, D_{ir_i})^T$ with $D_{ij} = D_i(t_{ij})$ are independent conditional on $\mathbf{\Lambda}_i$ we can write the likelihood for subject i as :

$$\begin{aligned}\mathcal{L}_i(\theta; \mathbf{X}_i, \mathbf{D}_i) &= \mathcal{L}_i(\theta; \mathbf{X}_i) \times \mathcal{L}_i(\theta; \mathbf{D}_i | \mathbf{X}_i) \\ &= \mathcal{L}_i(\theta; \mathbf{X}_i) \int \mathcal{L}_i(\theta; \mathbf{D}_i | \mathbf{\Lambda}_i = \lambda_i) \, dP^{\theta}_{\mathbf{\Lambda}_i | \mathbf{X}_i}(\lambda_i) \quad (5)\end{aligned}$$

where $\mathbf{\Lambda}_i = (\Lambda_{i1}, \cdots, \Lambda_{ir_i})^T$, with $\Lambda_{ij} = \Lambda_i(t_{ij})$, and where the set of parameters is denoted by $\theta = (\alpha, \beta, \gamma_0, \gamma_1, \eta, \Gamma, \sigma_e^2)$. We first evaluate the term $\mathcal{L}_i(\theta; \mathbf{D}_i | \mathbf{\Lambda}_i = \lambda_i)$, denoted more briefly $\mathcal{L}_{\mathbf{D}_i | \lambda_i}$ (here λ_i is a dummy integration variable belonging to $\mathbb{R}^{r_i}$). If $D_{im_i} = 1$, the disease occured in $(t_{i,m_i-1}, t_{im_i}]$, that is, the latent process $\Lambda_i(t)$ never attained the value η in $[0, t_{i,m_i-1}]$, and attained it in $(t_{i,m_i-1}, t_{im_i}]$. If $D_{ir_i} = 0$, the latent process never reached η during the study, that is in $[0, t_{ir_i}]$.

Hence we have to compute the following probabilities:

$$\begin{cases} \mathrm{P}[\min\limits_{t \in [0, t_{i,m_i-1}]} \Lambda_i(t) > \eta, \min\limits_{t \in (t_{i,m_i-1}, t_{im_i}]} \Lambda_i(t) \leq \eta | \mathbf{\Lambda}_i = \lambda] \text{ if } D_i(t_{ir_i}) = 1 \\ \mathrm{P}[\min\limits_{t \in [0, t_{ir_i}]} \Lambda_i(t) > \eta | \mathbf{\Lambda}_i = \lambda] \quad \text{if} \quad D_i(t_{ir_i}) = 0 \end{cases}$$

It can be shown (see appendix) that this part of the conditional likelihood

is as follows:

$$\mathcal{L}_{\mathbf{D}_i|\lambda_i} =$$
$$\left\{ \prod_{j=1}^{m_i-1} \left(1 - e^{\frac{-2(\lambda_{i,j-1}-\eta)(\lambda_{ij}-\eta)}{t_{ij}-t_{i,j-1}}}\right) I_{\{\lambda_{ij}>\eta\}} \times \left(e^{\frac{-2(\lambda_{i,m_i-1}-\eta)(\lambda_{im_i}-\eta)}{t_{im_i}-t_{i,m_i-1}}} \right)^{1-I_{\{\lambda_{im_i}<\eta\}}} \right\}^{D_{ir_i}}$$
$$\times \left\{ \prod_{j=1}^{r_i} (1 - e^{\frac{-2(\lambda_{i,j-1}-\eta)(\lambda_{ij}-\eta)}{t_{ij}-t_{i,j-1}}}) I_{\{\lambda_{ij}>\eta\}} \right\}^{1-D_{ir_i}} \quad (6)$$

In formula (5), the expectation of $\mathcal{L}_{\mathbf{D}_i|\mathbf{\Lambda}_i}$ with respect to the conditional density function of the latent process vector, $\mathbf{\Lambda}_i$, given $\mathbf{X}_i$ has to be taken. From a result of Lindley (1971) the distribution of $\mathbf{\Lambda}_i$, given $\mathbf{X}_i$ is a multivariate normal distribution of dimension r_i, with expected value:

$$\mathbf{M}_i = \mathbf{V}_i \frac{\gamma_1}{\sigma_e^2} \left[\mathbf{X}_i - \gamma_0 \mathbf{1} - \gamma_1(\mathbf{Z}_i^1 \alpha + \mathbf{Z}_i^3 \beta \mathbf{t}_i)\right] + (\mathbf{Z}_i^1 \alpha + \mathbf{Z}_i^3 \beta \mathbf{t}_i).$$

where $\mathbf{V}_i = (\frac{\gamma_1^2}{\sigma_e^2} I + \mathbf{G}_i^{-1})^{-1}$ is the variance of this distribution, with

$$\mathbf{G}_i = \mathbf{\Sigma}_i + \left(\mathbf{1}_{r_i} \mathbf{Z}_i^2 \quad , \mathbf{S}_i \mathbf{1}_{r_i} \mathbf{Z}_i^4 \right) \begin{pmatrix} \mathbf{\Gamma_a} & \mathbf{\Gamma_{ab}} \\ \mathbf{\Gamma_{ba}} & \mathbf{\Gamma_b} \end{pmatrix} \left((\mathbf{1}_{r_i} \mathbf{Z}_i^2)^T \quad , (\mathbf{1}_{r_i} \mathbf{Z}_i^4)^T \mathbf{S}_i \right),$$

where $\mathbf{S}_i = \text{Diag}(t_{i1}, \cdots, t_{ir_i})$, where element jk of $\mathbf{\Sigma}_i$ is equal to $\min(t_{ij}, t_{ik})$.

For the second part of the likelihood function, $\mathcal{L}_i(\theta; \mathbf{X}_i)$ denoted more briefly $\mathcal{L}_{\mathbf{X}_i}$, $\mathbf{X}_i$ has a normal distribution with mean vector $\gamma_0 \mathbf{1}_{r_i} + \gamma_1(\mathbf{Z}_i^1 \alpha + \mathbf{Z}_i^3 \beta \mathbf{t}_i)$ and variance matrix $\gamma_1^2 \mathbf{G}_i + \sigma_e^2 \mathbf{I}_{r_i}$.

If D_i is observed in continuous time the same approach can be taken for the likelihood construction, but since observing D_i (for uncensored subject

i) is observing T_η^i, formula (6) must be replaced by:

$$\mathcal{L}_{\mathbf{D}_i|\lambda_i} = \left\{ \prod_{j=1}^{m_i-1} \left(1 - e^{\frac{-2(\lambda_{i,j-1}-\eta)(\lambda_{ij}-\eta)}{t_{ij}-t_{i,j-1}}} \right) I_{\{\lambda_{ij}>\eta\}} \times f(T_\eta^i, \mu_i, \lambda_{i,m_i-1}, \eta) \right\}^{D_{ir_i}} \times \left\{ \prod_{j=1}^{r_i} (1 - e^{\frac{-2(\lambda_{i,j-1}-\eta)(\lambda_{ij}-\eta)}{t_{ij}-t_{i,j-1}}}) I_{\{\lambda_{ij}>\eta\}} \right\}^{1-D_{ir_i}},$$

with $f(.)$ as in formula (3), and $\mu_i = \mathbf{Z}_i^3 \beta + \mathbf{Z}_i^4 b_i$.

3.2 Selection of the sample

The sample may be selected according to a condition C_i, so we must use a conditional likelihood $\frac{\mathcal{L}_i(\theta)}{\mathrm{P}^\theta(C_i)}$. It may often be the case that subjects having experienced the failure before the first observation are excluded from the sample; the condition is $D_i(t_{i1}) = 0$. So we have to compute the probability that the minimum of the latent process Λ_i in $[0, t_{i1}]$ be greater than η

$$\mathrm{P}(D_{i1} = 0) = \mathrm{P}(\psi_i(\mathbf{a}_i) > \eta) - \int\int F(t_{i1}, \mu_i(\mathbf{b}_i), \psi_i(\mathbf{a}_i), \eta) dP_{\mathbf{a}_i,\mathbf{b}_i},$$

where $F(.)$ is as formula (4) and $dP_{\mathbf{a}_i,\mathbf{b}_i}$ represents as in formula (2) the integration relatively to the distribution of the random effects conditionally on $\psi_i(\mathbf{a}_i) > \eta$.

4 Algorithm for maximising the likelihood

4.1 Pseudo-likelihood

The procedure of maximisation of the likelihood is relatively slow because of the numerical integration. So to find the initial values, we construct a pseudo-likelihood, for which we use the information of indirect measurements, that

is, $\mathcal{L}_{\mathbf{X}_i}$ as in section (3.1), for the likelihood pretending we observe $D_{ir_i} = 1$ with t_{ir_i} fixed and D_i independent of X_i. Thus the pseudo-likelihood is:

$$\mathcal{L}_{\mathbf{X}_i} \times \mathrm{P}(D_{ir_i} = 1)^{D_{ir_i}} \times \mathrm{P}(D_{ir_i} = 0)^{1-D_{ir_i}}.$$

Ignoring the expectation with respect to the density of the random effects, $\mathrm{P}(D_{ir_i} = 1)$ can be approximated by $F(t_{ir_i}, \mu_i(\beta), \psi_i(\alpha), \eta)$, where $\psi_i(\alpha) = \mathbf{Z}_i^1\,\alpha$ and $\mu_i(\beta) = \mathbf{Z}_i^3\,\beta$, with $F(.)$ as in formula (4). This pseudo-likelihood is easily maximised and provides good initial values.

4.2 Algorithm of maximisation

We maximise the likelihood using a modified version of the Newton-Raphson algorithm, proposed by Marquardt (1963). The computation of the likelihood involves a multiple numerical integration for each subject and we must also compute the first and second derivatives, which we do by numerical differentiation. Multiple integrations are usually done by Monte-Carlo algorithms (see Evans and Swartz, 2000). The simplest one is to approximate $\mathcal{L}_{D_i|X_i}$ by $\frac{1}{N}\sum_{k=1}^{N}\mathcal{L}_{\mathbf{D}_i|\mathbf{\Lambda}_i}(\lambda_k)$, where λ_k is generated from the conditional distribution of Λ_i given X_i.

5 Simulation study

In the simulation study we consider a model with one explanatory variable and a random intercept and a random slope; so the model is as follows:

$$\Lambda_i(t) = Z_i\,\alpha + a_i + (\beta_0 + \beta_1\,Z_i + b_i)\,t + W_i(t)$$

and

$$X_{ij} = \gamma_0 + \gamma_1\,\Lambda_i(t_{ij}) + \varepsilon_{ij}$$

and $D_i(t) = I_{\{T_\eta^i \leq t\}}$. We have to estimate the set of parameters $\{\alpha,\ \sigma_a^2,\ \sigma_b^2,\ \sigma_{ab},\ \sigma_e^2,\ \beta_0,\ \beta_1,\ \gamma_0,\ \gamma_1,\ \eta\}$.

Marginally, the model of X_i is:

$$X_{ij} = \gamma_0 + Z_i\,\alpha\,\gamma_1 + a_i\,\gamma_1 + \gamma_1\,(\beta_0 + Z_i\,\beta_1 + b_i)\,t_{ij} + \gamma_1\,W_i(t_{ij}) + \varepsilon_{ij}.$$

We may base inference on observations of the X_{ij} only, using what would be a marginal likelihood relative to the full model. It is interesting to compare the precision obtained with this marginal inference based only on the observation of X_i (ignoring the information on D_i) to that obtained using the joint model. Moreover we apply in this simulation the same artificial missing data mechanism as we apply in the example: observations of X_i after t_{im_i} (when dementia has been diagnosed) are ignored. This leads to non-informative missing data in the joint model because the fact that data are missing depends on what has been observed. The missing data mechanism will be informative in the marginal model since D_i is not modeled; this happens because the distribution of $(X_{i1}, \ldots, X_{im_i})$ given $m_i = c$ is not the same as the distribution of $(X_{i1}, \ldots, X_{ic})$ for fixed c.

In this study, 200 replications were done and in each of them we simulated 200 subjects. The explanatory variable Z_i was simulated according to the Bernoulli distribution with parameter 0.25. The first visit time of each subject was at the age 65 years plus a number which has been chosen randomly from the set $\{1, 2, \cdots, 8\}$; there were 5 years between all visits. As explained above, observation were stopped at t_{im_i} for subject with observed event. The number of visits was less than or equal to 6, and distributed as in Table 1. Both models were simulated using the values for parameters in Table 2 called true parameters. For the computation of each integral, 2000 replications have been used. The algorithm generally converged well (convergence being judged on difference in the loglikelihoods, distance in the parameter

space and most importantly, norm of the gradient) in about 10 iterations and took about 25 minutes, on a PC (pentium 3); in the simulation study it failed to converge in about 5% of the cases, probably due to some extreme realized values of the observed variables.

The results are summarized in Table 2. The estimated slope in the marginal model (β_0) seems to have a bias which can be explained by the informative censoring described above. The joint model yields smaller standard deviations for all the estimators; in addition it allows to estimate η and to make inference on the failure. We may note in addition that the correlation between the random effects seems difficult to estimate, at least with this sample size.

6 Application

The proposed approach was applied to the joint modeling of dementia and a psychometric test, the Mini Mental State (MMS) (Folstein et al. 1975), using the data of the PAQUID cohort.

The PAQUID program on cerebral aging is based on a large cohort randomly selected in a population of subjects aged 65 years or more, living at home in two departments of southwest France (Gironde and Dordogne). There were 3675 subjects non demented at entry in the cohort considered and each subject has been visited six times or less, between 1988 and 2000; 428 cases of incident dementia were observed. At each visit the MMS was measured and diagnosis of dementia was made by neurologists based on the NINCDS-ADRDA criteria (for details see Letenneur et al., 1999). The information given by the diagnosis of dementia on the latent process is considered as independent from the MMS value.

The risk of developing dementia was modeled as a function of age, so the data were left-truncated and the truncation variable was the age at entry in the cohort (for details see Commenges et al., 1998). There was also the problem of determining the origin of time: it would not make sense to model the decline of cognitive ability as having a linear trend from birth of the subject. Rather we made the assumption of a linear trend from an age located between 50 and 65. We used profile likelihood to determine this origin of time, that is, for each given value of the starting age we ran the program and computed the likelihood; then we chose the value which maximised the likelihood. This starting age was determined as 58 years.

In principle an illness-death model would be appropriate (Joly et al., 2002) if one did not want to jointly model dementia and cognitive performance; however approximate inference for the transition towards dementia can be made by considering death as censoring. It is likely that this produces a bias but it is not possible to treat the joint model rigorously with existing models.

In this application of our model, the latent process represents cognitive ability; if it goes under a threshold η, dementia occurs; cognitive ability is indirectly measured by a psychometric test, the MMS. Since the distribution of the MMS is far from normal we use the transformation $\sqrt{30 - MMS}$, already used by Jacqmin-Gadda et al. (1997). When dementia was diagnosed, further observations of the MMS were not taken into account, as already explained in the simulation section. This was done because the behaviour of the latent process may be different for non-demented and demented subjects.

We entered gender as an explanatory variable, which will be denoted by SEX=1 if female and SEX=0 if male. A model with a random effect in slope was tried, but the variance of the random effect was estimated to be zero.

Thus the estimated model contains only a random intercept and is as follows:

$$\Lambda_i(t) = \alpha\,\mathrm{SEX}_i + a_i + (\beta_0 + \beta_1\,\mathrm{SEX}_i)\,t + W_i(t)$$

where $a_i \sim N(0, \sigma_a^2)$. The estimated threshold was $\hat{\eta} = -12.38$, and

$$X_{ij} = \gamma_0 + \gamma_1\,\Lambda_i(t_{ij}) + \varepsilon_{ij}$$

The results are summarized in Table 3. With the estimated values of the parameters γ_0, γ_1 and η the mean value of X at onset of dementia is 3.2 corresponding to a MMS value of 19.7. This is compatible with clinical practice: although the diagnosis of dementia is not based directly on the MMS, most diagnosed subjects have a MMS below 24, a threshold conventionally used to define cognitive impairment; on the other hand not all subjects scoring below 24 are diagnosed as demented.

The hazard function of the latent process model (formula 2), for men and women, is shown in figure 1. As expected there is a negative slope for the latent process ($\hat{\beta}_0 = -0.216$). Women start higher than men ($\hat{\alpha} = 0.712$) but have a steeper slope ($\hat{\beta}_1 = -0.09$); all these effects are significantly different from zero. This is in accordance with previous findings (Commenges et al., 1998; Joly et al., 2002) that women in the older age group have a higher risk of dementia than men. There is also a random effect on the intercept with variance significantly different from zero.

7 Conclusion

We have developed a joint modeling of a marker and an event which fits well in situations where the event can be conceptualized as defined by the crossing of a barrier by a given process; this is the case of dementia and this gives a better interpretability of the parameters. Moreover our work shows that it

is possible to maximize a likelihood using Newton-Raphson type algorithm even when the likelihood involves many numerical integrals.

One limit of the model is the assumption of a linear trend in time of the latent process which allows using relatively simple formulas available for Brownian motion with (linear) drift. Removing this assumption would make the computation much heavier. Thus it would be useful to develop diagnosis tests for this model.

An interesting possibility would be to estimate the values of the random effects and of the Brownian motion for a subject i given the observations; also prediction could directly be done by estimating the conditional probability given the observations of being demented at a given time after the last visit. This could in principle be done by an empirical Bayes approach, computing conditional expectations given the observations. However, in view of the complexity of the model this would have to be done by simulation.

Further extensions would involve both multiple markers and events. For instance in the application to cognitive decline, several psychometric tests may be performed and it would be useful to model death as well as dementia.

Appendix: Calculation of the likelihood

To compute the first part of the likelihood function denoted by $\mathcal{L}_{\mathbf{D}_i|\mathbf{\Lambda}_i}$ we evaluate the following probabilities

$$\begin{cases} \mathrm{P}[\min\limits_{t\in[t_{i0},t_{i,m_i-1}]}\Lambda_i(t) > \eta, \min\limits_{t\in[t_{i,m_i-1},t_{im_i}]}\Lambda_i(t) \leq \eta|\mathbf{\Lambda}_i=\lambda_i] \text{ if } D_i(t_{im_i})=1 \\ \mathrm{P}[\min\limits_{t\in[t_{i0},t_{ir_i}]}\Lambda_i(t) > \eta|\mathbf{\Lambda}_i=\lambda_i] \text{ if } D_i(t_{ir_i})=0 \end{cases}$$

Now we partition the interval $[t_{i0}, t_{im_i}]$ to $[t_{i0}, t_{i1}], \cdots, [t_{i,m_i-1}, t_{im_i}]$, and in the same way the interval $[t_{i0}, t_{ir_i}]$ to $[t_{i0}, t_{i1}], \cdots, [t_{i,r_i-1}, t_{ir_i}]$ where $t_{i0}, \cdots, t_{ir_i}$ are fixed, we consider independent Brownian bridges obtained by conditioning Λ by its values at the bounds of these intervals. Thus minimum of the process in $[t_{i,j-1}, t_{ij}]$ depends only on $\Lambda_{i,j-1}, \Lambda_{ij}$, so $\mathcal{L}_{D_i|\Lambda_i}$ can be written as

$$\mathcal{L}_{D_i|\Lambda_i} = \{\prod_{j=1}^{m_i-1} \mathrm{P}[\min_{s\in[0,\Delta t_{ij}]}(\Lambda_i(s+t_{i,j-1})-\lambda_{i,j-1}) > \eta-\lambda_{i,j-1}|\Delta\Lambda_{ij}=\Delta\lambda_{ij}]$$
$$\times \mathrm{P}[\min_{s\in[0,\Delta t_{im_i}]}(\Lambda_i(s+t_{i,m_i-1})-\lambda_{i,m_i-1}) \leq \eta-\lambda_{i,m_i-1}|\Delta\Lambda_{im_i}=\Delta\lambda_{im_i}]\}^{D_{im_i}}$$
$$\times \{\prod_{j=1}^{r_i} \mathrm{P}[\min_{s\in[0,\Delta t_{ij}]}(\Lambda_i(s+t_{i,j-1})-\lambda_{i,j-1}) > \eta-\lambda_{i,j-1}|\Delta\Lambda_{ij}=\Delta\lambda_{ij}]\}^{1-D_{ir_i}}$$

where $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ and $\Delta\Lambda_{ij} = \Lambda_{ij} - \Lambda_{i,j-1}$. $\Lambda_i(s+t_{i,j-1}) - \Lambda_{i,j-1}$ conditioned on $\Delta\Lambda_{ij} = \Delta\lambda_{ij}$ has same law as a Brownian motion conditioned on $B(\Delta t_{ij}) = \Delta\lambda_{ij}$ (this is a Brownian bridge starting at zero and ending at $\Delta\lambda_{ij}$, in particular it is interesting to see that the law of this process does not depend on the linear drift). Thus using standard results (Karatzas and Shreve, 1991; Klebaner, 1998) on the distribution of the minimum of a Brownian bridge we obtain formula (6).

References

O.O. Aalen, H.K. Gjessing, Understanding the shape of the hazard rate: a process point of view. *Statistical Science* vol. 16 pp. 1-22, 2001.

D. Commenges, L. Letenneur, P. Joly, A. Alioum and J. F. Dartigues, Modelling age-specific risk: application to dementia. *Statistics in Medicine* vol. 17 pp. 1973-1988, 1998.

D.R. Cox, Some Remarks on Failure-times, Surrogate markers, degradation, wear, and the quality of life. *Lifetime Data Analysis* vol. 5 pp. 307-314, 1999.

M. Evans, T. Swartz, Approximating integrals via Monte Carlo and deterministic methods. *Oxford University Press; Oxford* 2000.

P. Diggle, An approach to the analysis of repeated measurements. *Biometrics* vol. 44 pp. 959-971, 1988.

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

M.F. Folstein, S.E. Folstein, P.R. McHugh. Mini-Mental State. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* vol. 12 pp. 189-98, 1975.

R. Henderson, P. Diggle, A. Dobson, Joint modeling of longitudinal measurements and event time data. *Biostatistics* vol. 1 pp. 465-480, 2000.

J.W. Hogan, N.M. Laird, Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* vol. 16 pp. 259-272, 1997.

H. Jacqmin-Gadda, C. Fabrigoule, D. Commenges, JF. Dartigues. A 5-year longitudinal study of the mini-mental state examination in normal aging. *American Journal of Epidemiology* vol. 145 pp. 498-506 1997.

P. Joly, D. Commenges, C. Helmer, L. Letenneur, A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* vol 3 pp. 433-443, 2002.

I. Karatzas, S.E. Shreve, Brownian motion and stochastic calculus. -2nd ed. *Springer-Verlag, New-York* 1991.

F.C. Klebaner, Introduction to stochastic calculus with applications. *Imperial College Press, London* 1998.

M.L.T. Lee, V. DeGruttola, D. Schoenfeld, A model for markers and latent health status. *Journal of Royal Statistical Society: Series B* vol. 62 pp. 747-762, 2000.

L. Letenneur, V. Gilleron, D. Commenges, C. Helmer, J.M. Orgogozo, J.F. Dartigues, Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the Paquid project. *Journal of Neurology Neurosurgery and Psychiatry* vol. 6 pp. 177-183, 1999.

D.V. Lindley, Bayesian statistics, A Review. *SIAM* Philadelphia, PA. 1971.

D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics* pp. 431-441, 1963.

G.A. Whitmore, M.J. Crowder, J.F. Lawless, Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Analysis* vol. 4 pp. 229-251, 1998.

M. S. Wulfsohn, A. A. Tsiatis, A joint model for survival and longitudinal data measured with error. *Biometrics* vol. 53 pp. 330-339, 1997.

Table 1: Distribution of number of visits in the simulation study

Probability	0.05	0.15	0.2	0.2	0.2	0.2
Num. of visits	1	2	3	4	5	6

Table 2: Results of the simulation study with 200 replications, of 200 subjects. For each parameter: first row: linear model; second row: joint model

Parameter in LMM[1] in JMLP[2]	True value	Average of $(\hat{\theta})$	empirical S.D.$(\hat{\theta})$	Average of $\widehat{S.E.(\hat{\theta})}$
σ_e	0.44	0.430	0.0432	0.0402
		0.447	0.0138	0.0173
σ_a	2.	2.015	0.929	0.919
		2.053	0.709	0.724
σ_b	0.2	0.264	0.141	0.129
		0.253	0.0804	0.0838
σ_{ab}	0.1	0.063	0.312	0.384
		0.057	0.249	0.293
α	-2.	-2.178	1.009	0.987
		-2.246	0.935	0.829
β_0	-0.4	-0.238	0.0914	0.0823
		-0.459	0.0737	0.0693
β_1	-0.2	-0.178	0.106	0.102
		-0.223	0.0856	0.0871
γ_0	1.2	1.219	0.0625	0.0651
		1.193	0.0618	0.0613
γ_1	-0.19	-0.199	0.0424	0.0382
		-0.178	0.0183	0.0224
η	-11.	$--$	$--$	$--$
		-11.492	1.191	1.354

[1]: Linear mixed model, [2]: Joint model with latent process

Table 3: Estimated parameters of the joint model for MMS and dementia and their standard deviations (the PAQUID study $N = 3675$)

Parameter	Estimation	$\widehat{S.D(\hat{\theta})}$
σ_e	0.514	0.004
σ_a	1.405	0.302
α	0.712	0.312
β_0	-0.216	0.014
β_1	-0.090	0.017
γ_0	1.116	0.04
γ_1	-0.169	0.004
η	-12.38	0.328

Figure 1: Estimated hazard function of dementia in the latent process model by gender (PAQUID study)