A joint model for longitudinal and survival data based on a continuous-time latent Markov model

Un modello congiunto per dati longitudinali e di sopravvivenza basato su un processo Markoviano latente in tempo continuo

Alessio Farcomeni and Francesco Bartolucci

Abstract A shared-parameter approach for jointly modeling longitudinal and survival data is proposed, which allows for time-varying random effects that enter both in the longitudinal and survival processes. The distribution of these random effects is modeled according to a continuous-time hidden Markov chain, so that latent transitions may occur at any time point. Our formulation allows for: (*i*) informative drop-out with precise time-to-event outcomes, while existing approaches are all based on drop-out at longitudinal measurement times and (*ii*) completely non-parametric treatment of unequally spaced intervals between consecutive measurement occasions (even not in the presence of drop-out). For maximum likelihood estimation we propose an algorithm based on coarsening. The resulting estimator is studied by simulation. The approach is illustrated by an application to data about patients suffering from mildly dilated cardiomyopathy.

Abstract *Si propone un modello congiunto per dati longitudinali e di sopravvivenza. La distribuzione degli effetti casuali condivisi è modellata sulla base di un processo latente in tempo continuo e spazio degli stati discreto. La formulazione proposta permette di prevedere (i) drop-out informativo in presenza di tempo-a-evento precisamente misurato, mentre gli approcci attualmente disponibili sono basati su indicatori al momento della misurazione del dato longitudinale, e (ii) un trattamento non-parametrico per il caso di intervalli non-omogenei tra occasioni di misura, anche in assenza di drop-out. Per la stima di massima verosimiglianza sviluppiamo un algoritmo basato sul coarsening. Lo stimatore risultate é valutato per simulazione. L'approccio é illustrato tramite una applicazione a dati su pazienti cardiomiopatia dilatativa di grado lieve.*

Key words: Baum-Welch recursion, Informative drop-out, Unequally spaced times

A. Farcomeni
Sapienza - University of Rome, e-mail: alessio.farcomeni@uniroma1.it

F. Bartolucci
University of Perugia e-mail: francesco.bartolucci@unipg.it

1 Introduction

Informative drop-out in longitudinal studies is often treated by linking a model for time to drop-out and one for the longitudinal outcome. Many models devised for informative drop-out assume that subject-specific parameters are time-constant. This is a limitation as unobserved factors affecting the outcomes and the relationship between longitudinal and survival outcomes might evolve over time in an unpredictable way, especially when the follow-up is relatively long. One exception is [1], who propose a discrete-time event-history approach based on latent Markov models [2], which naturally accommodate time-dependent unobserved heterogeneity. Nevertheless, the approach in [1] has two limitations: (*i*) the event-history component models drop-out, by a conditional logit model, as occurring within a time interval, therefore ignoring precise follow-up time information; (*ii*) latent transitions, as common in latent Markov models, are based on a discrete-time stochastic process and hence transitions may only occur at visit times. In terms of interpretation, assuming that transitions may occur only at certain time occasions is rather unrealistic and the explicit use of an hazard function is preferable to that of a conditional logit model.

In order to overcome the above limitations, we propose a shared-parameter model characterized by the following features. First of all the time-varying unobserved heterogeneity is accounted for by a continuous-time discrete-state hidden Markov model [5], parameterized by an initial probability vector and an infinitesimal transition matrix. In this respect our approach can be seen as a complete generalization of the relevant work by [3], which is limited to $k = 2$ and to missing-at-random data. Second, for the survival time we assume a Weibull model with hazard function depending on the (entire) trajectory of the continuous-time latent variable. A latent class model (with time-constant subject-specific parameters) is obtained whenever the infinitesimal transition matrix is constrained to have all elements equal to 0.

For model fitting we introduce a novel method that provides maximum likelihood estimates. The approach is based on a time discretization, in a certain number of windows of arbitrary length, and on an extension of the Baum-Welch recursions. It converges in an accurate and stable way. This algorithm also represents an advance in the literature about estimation of continuous hidden Markov models in general, with respect to computational demands, ease of implementation, and stability.

In the following we illustrate in some detail the assumptions of the proposed model, then we describe the approach to parameter estimation and, finally, we outline an application based on data about patients suffering from Mildly Dilated CardioMyopathy (MDCM).

2 Shared-parameter continuous-time latent Markov and survival models

Consider a sample of n individuals and for individual i, with $i = 1,\ldots,n$, let $T_i = \min(T_i^*, C_i)$ be the survival time taken as the minimum between the true event time T_i^* and the censoring time C_i. Furthermore, let Δ_i be the corresponding event indicator defined by $\Delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function equal to 1 if its argument is true and to 0 otherwise. The outcome $Y_i(t)$, which arises from a natural exponential family, is repeatedly observed at arbitrary time points t_{ij}, $j = 1,\ldots,j_i$, where j_i is the number of observations; also let $Y_{ij} = Y_i(t_{ij})$. We assume that the longitudinal process is associated with T_i^*, namely with the true event time, but, as customary in survival analysis, is independent of the censoring time C_i. In general, realizations of random variables are denoted by small letters, so that, for instance, t_i is the observed value of T_i and δ_i is the observed value of Δ_i.

We denote by w_i a row vector of (time-fixed) baseline covariates to be used in modeling the survival process. For the longitudinal process, we denote by $x_i(t)$ a vector of predictors at time t and we also let $x_{ij} = x_i(t_{ij})$, $j = 1,\ldots,j_i$.

The proposed model is based on two equations. Specifically, the model for the longitudinal outcome is formulated along the usual lines of mixed-effects models and the model for the time-to-event outcome is based on a subject-specific hazard function as in Cox-type models. More formally, we assume that

$$\begin{aligned} g(\mu_{ij}) &= \alpha_i(t_{ij}) + x_{ij}'\beta, \quad j = 1,\ldots,j_i, \\ h(t_i^*) &= h_0(t_i^*)\exp\{\alpha_i(t_i^*)\phi + w_i'\psi\}, \end{aligned}$$

where $g(\cdot)$ is a link function of the conditional expectation of Y_{ij} denoted by μ_{ij} and $h(\cdot)$ is the hazard function, with $h_0(\cdot)$ being a baseline hazard. In this paper we will assume a Weibull parametric form for $h_0(\cdot)$, that is, $h_0(t) = \nu t^{\nu-1}$, resulting in an Accelerated Failure Time (AFT) model for the survival part. Other parametric choices, or even a non-parametric specification, are possible. We assume that $\alpha_i(t)$ follows a time-continuous Markov process, whereas β and ψ are fixed parameter vectors for the covariates, and ϕ is a parameter for the effect of the latent process on the survival process. Note that several generalizations, including the case of more than one parameter being time-dependent according to the latent process, are straightforward. Regarding the distribution of Y_{ij}, our model has the same degree of flexibility as generalized linear models. It is worth also stressing that the hazard function depends on the *entire* trajectory of the random effect $\alpha_i(t)$, and not only on $\alpha_i(t_{ij})$.

Unlike usual formulations, random intercepts are assumed to be time varying. This greatly enhances model flexibility. In particular, as already mentioned, we assume that the random effects follow a continuous-time (discrete-state) Markov chain [5], with state-space $\{\xi_1,\ldots,\xi_k\}$ having elements collected in the column vector ξ. We assume that the transition function of the latent chain satisfies the Chapman-Kolmogorov equations, and specify its Q-matrix based on positive off-diagonal el-

ements q_{uv} for $u,v = 1,\ldots,k$ and $v \neq u$. By definition, the diagonal elements are given by $-q_u$, with $q_u = \sum_{v=1,v\neq u}^{k} q_{uv}$,$u = 1,\ldots,k$. Accordingly, for the longitudinal outcome transitions from time t to time $t+s$ are collected in the $k \times k$ matrix $\Pi_s = e^{sQ}$, where e denotes the *matrix exponential* operator, that is, $e^{sQ} = \sum_{n=0}^{\infty} \frac{s^n Q^n}{n!}$. Note that irregularly spaced time occasions are directly accommodated, and hence our model also generalizes [3] simply by restricting it to the first equation. We also define the *jump matrix* R as a matrix with off-diagonal elements $r_{uv} = q_{uv}/q_u$, and collect initial probabilities π_u in the column vector π.

The latent process captures the time-varying unobserved heterogeneity linking the longitudinal and survival outcomes. The shared-parameter formulation is in the spirit of copula models [6]. We also recall that, according to this process, the sojourn time in each state u has an exponential distribution with parameter q_u, denoted as $\text{Exp}(q_u)$, whereas the probability of moving at the end of the sojourn time to state v is equal to the suitable element of the jump matrix R.

3 Estimation

It is straightforward to check that the complete likelihood of our proposed model depends on the entire trajectory of the continuous-time latent process, through the integrals involved in the time-to-event component. This makes it hard to efficiently compute the observed likelihood: the classical Baum-Welch recursion is not directly applicable, even after their extension to continuous time processes, due to lack of certain conditional independence statements. To derive inference we build a sequence of equally spaced windows corresponding to time fixed points $\bar{t}_1,\ldots,\bar{t}_M$, with $\bar{t}_1 = 0$. The first window spans the time interval from $\bar{t}_1$ to $\bar{t}_2$, the second from $\bar{t}_2$ to $\bar{t}_3$, and so on. These time points are chosen so that each observation time t_{ij} corresponds to one of them. Let $\bar{y}_{im}$ denote the observation at time $\bar{t}_m$ for individual i, which may be missing for certain time occasions, $\bar{x}_{im}$ be the corresponding vector of covariates, and $\bar{U}_{im}$ the corresponding latent variable. Also let $\bar{y}_{i,\leq m}$ be vector of observations available until time $\bar{t}_m$.

The following forward recursion can now be used. Consider the joint density $f_{im}(u) = f(\bar{y}_{i,\leq m}, t_i \geq \bar{t}_m, \bar{U}_{im} = u)$ referred the observation availably until time $\bar{t}_m$ for individual i, latent state at the same time occasion, and for the event that the individual survives time $\bar{t}_m$. We have that $f_{i1}(u) = \pi_u f(\bar{y}_{i1}|\bar{U}_{i1} = u), \quad u = 1,\ldots,k$, and $f_{im}(v) = f(\bar{y}_{im}|\bar{U}_{i1} = v)\sum_{u=1}^{k} \pi_{v|u} f_{i,m-1}(u) S_{m-1}(\bar{t}_m, u), \quad m = 1,\ldots,m_i,\, v = 1,\ldots,k$, where, in general, we have $S_m(\bar{t},u) = \exp\{-H_m(\bar{t},u)\}$ with

$$H_m(\bar{t},u) = \int_{\bar{t}_m}^{\bar{t}} \exp(\xi_u\phi + w_i'\psi)\nu t^{\nu-1} dt = \exp(\xi_u\phi + w_i'\psi)(\bar{t}^{\nu} - \bar{t}_m^{\nu}), \quad u = 1,\ldots,k,$$

$f(\bar{y}_{im}|\bar{U}_{i1} = v)$ is set equal to 1 if the observation is not available at time t_m, and m_i be largest value of m such that $\bar{t}_m \leq t_i$. For individual i we have the contribution to the likelihood given by $f(y_i,t_i,d_i) = \sum_{u=1}^{k} f_{im_i}(u) h(t_i)^{\delta_i} S_{m_i}(u,t_i)$. Regarding the

transition probabilities $\pi_{v|u}$, note that these are the elements of the $k \times k$ matrix Π_a obtained as $\exp(aQ)$, where $a = \bar{t}_{m+1} - \bar{t}_m$ is the length of each time window.

The log-likelihood function to be maximized is then $\ell(\theta) = \sum_{i=1}^{n} \log f(y_i, t_i, d_i)$. In order to maximize this function we also need a backward recursion. In particular, let $g_{im}(u) = f(\bar{y}_{i,>m}, t_i, \delta_i | t_i > t_m, U_{im} = u)$. For $m = m_i$ have that $g_{im_i}(u) = h(t_i)^{\delta_i} S_{m_i}(t_i, u)$ and $g_{im}(u) = S_m(\bar{t}_{m+1}, u) \sum_{v=1}^{k} \pi_{v|u} g_{i,m+1}(v) f(\bar{y}_{i,m+1}, U_{i,m+1} = v)$ for $m < m_i$. From this recursion we can obtain two posterior distributions used to update the parameters π and Π_a. In particular, we have that

$$p(U_{im} = u | y_i, t_i, \delta_i) = \frac{f_{im}(u) g_{im}(u)}{f(y_i, t_i, d_i)}, \quad m = 1, \ldots, m_i,\ u = 1, \ldots, k.$$

Moreover, we have that

$$p(U_{im} = u, U_{i,m+1} = v | y_i, t_i, \delta_i) = \frac{f_{im}(u) S(u, \bar{t}_{m+1}) \pi_{uv} f(\bar{y}_{i,m+1}, U_{i,m+1} = v) g_{i,m+1}(v)}{f(y_i, t_i, d_i)},$$
$$m = 1, \ldots, m_i - 1,\ u, v = 1, \ldots, k.$$

Then, we update these parameters as $\pi_u = \frac{1}{n} \sum_{i=1}^{n} p(U_{i1} = u | y_i, t_i, \delta_i)$, $u = 1, \ldots, k$, and

$$\pi_{uv} = \frac{\sum_{i=1}^{n} \sum_{m=1}^{m_i - 1} p(U_{im} = u, U_{i,m+1} = v | y_i, t_i, \delta_i)}{\sum_{i=1}^{n} \sum_{m=1}^{m_i - 1} \sum_{\bar{v}=1}^{k} p(U_{im} = u, U_{i,m+1} = \bar{v} | y_i, t_i, \delta_i)}, \quad u, v = 1, \ldots, k.$$

The infinitesimal transition matrix Q is obtained from Π_a by inverting $\exp(aQ)$.

To update the other parameters, we explicit the expected value of the complete log-likelihood. In particular, regarding the third component about the survival process we have

$$\mathrm{E}\{\ell_3(\theta)\} = \sum_{i=1}^{n} \mathrm{E} \left\{ \delta_i \log h_0(t_i | U_{im_i}) - \sum_{m=2}^{m_i} H_{m-1}(\bar{t}_m, U_{i,m-1}) - H_m(t_i, U_{im_i}) \right\}.$$

Regarding the derivative of $\ell(\theta)$ with respect to the model parameters, let τ denote any of the elements of θ apart from those involved in the latent process; we apply the general rule

$$\ell(\theta) = \sum_{i=1}^{n} f(y_i, t_i, d_i)^{-1} \frac{\partial f(y_i, t_i, d_i)}{\partial \tau}.$$

4 Application to MDCM data

We illustrate now the proposed approach using an original study on a cohort of patients affected by MDCM, a primary myocardial disease characterized by left

ventricular systolic dysfunction and dilation. For more details on the pathology, see [4].

Prognostic measurements were taken at basal time for $n = 642$ patients, who were followed-up until urgent heart transplant or death occurred. There were 212 events during follow-up, which lasted up to 25 years. If censoring (administrative or due to the event) did not occur, measurement of longitudinal biomarkers were taken at visits scheduled at months 6, 12, 24, 48, 72, 96, and 120. Hence each patient has a maximum of 8 longitudinal measurements, with 79 patients having complete records.

The longitudinal outcome is the New York Hearth Association (NYHA) classification, a direct measure of discomfort caused by the disease. Specifically, for each subject at each follow-up occasion we measure an indicator of being in NYHA class III or IV, indicating the presence of strong limitations to physical activity, and the occurrence of dyspnea and discomfort during ordinary activities or even at rest. For the longitudinal model we parameterize probability of high NYHA class as a function of $t > 0$ (indicating medical treatment according to international guidelines), an indicator of history of heart disease in the family, and the left ventricular Ejection Fraction (EF). The latter is a measure of the proportion of blood that is pumped out of the left ventricle at each heart beat.

It is natural to expect that a continuous-time model is more appropriate for the data at hand than any model assuming latent transitions occurring at visit times. In fact, latent states shall be interpreted as patients' frailty beyond that summarized by the predictors, and changes in disease status (and hence propensity to event and/or change of NYHA class) obviously can occur at any time point and not necessarily on the day of the visit by the doctor. Further, a strong dependence between NYHA class and the event is expected, with patients in NYHA classes III or IV being at higher risk of death.

For interpretability reasons, EF has been centered at 30 (which is believed to be a significant threshold, where EF < 30 indicates patients at risk of heart failure).

For our model fitting procedure we evaluate several values for M, and end up fixing $M = 200$, which is well above values guaranteeing stability of estimates. Using the Bayesian information criterion we select $k = 3$. In order to estimate standard errors we perform a non-parametric bootstrap procedure based on $B = 1000$ replicates. In Table 1 we report parameter estimates for the manifest distribution, along with an indication of significance at the 5% level.

The estimate of Q is better understood after computation of the time-specific transition matrix. For this purpose, we report Figure 1 where the inhomogeneous transition matrix at each time t is reported.

The results indicate an important role of all predictors, with the exception of history of hearth disease for survival. Comparing $k = 1$ with $k > 1$, it is clear that taking into account unobserved heterogeneity leads to a more clear identification of the roles of EF and family history for NYHA classification. The effect of family history doubles when passing from $k = 1$ to $k = 3$, while the effect of each percentage point of EF is almost three times larger. Hence, based on our results, doctors should probably pay more attention to EF and family history than expected when assessing

Table 1 *MDCM data: parameter estimates for the manifest distribution, different values of k. An asterisk indicates statistical significance at the 5% level.*

		k			
	Effect	1	2	3	4
logit NYHA	ξ_1	-0.967*	-4.745*	-4.556*	-6.354*
	ξ_2	-	-0.164	-1.446*	-2.182*
	ξ_3	-	-	2.891*	-1.481*
	ξ_4	-	-	-	2.902*
	$t > 0$	1.047*	2.289*	0.920*	0.847*
	history	0.611*	0.724*	1.169*	1.126*
	EF	0.056*	0.094*	0.134*	0.137*
survival	ϕ	0.000	-0.475*	-0.314*	-0.310*
	history	-0.125	0.031	-0.019	-0.001
	EF	-0.058*	-0.048*	-0.049*	-0.058*
	ν	0.799*	0.781*	0.762*	0.789*

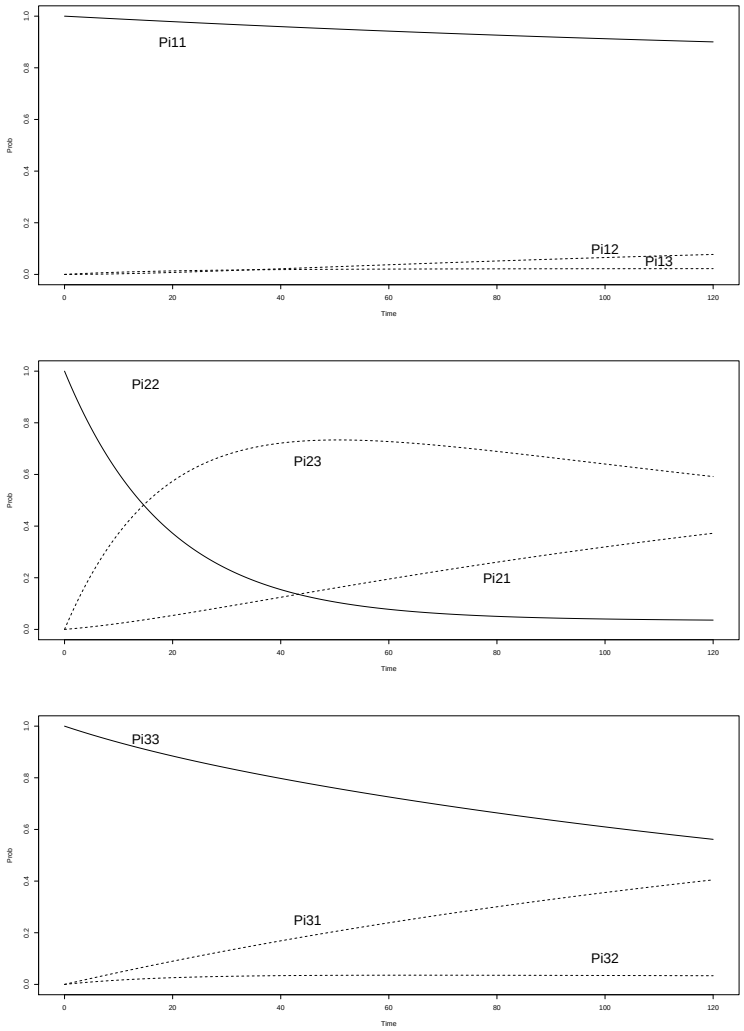
prognosis to high NYHA classes. The estimate of ϕ is negative and significant in all cases, indicating as expected that subjects with, for instance, dyspnea during ordinary activities are at higher risk of death than patients without clear signs of heart insufficiency.

When $k = 3$ three clearly separate groups of patients are identified. Even when they have the same history, EF and timing configuration, patients might be different due to unobserved factors. A group of patients (about 30% at baseline time) is at a very low risk. From Figure 1 it can be observed that this group of patients is slightly stable, with low probability of transitions to different states during follow-up. A second group (about 60%) is at slightly larger propensity to high MDCM at baseline. These patients are very likely to change state during follow-up, with many switchings to an even higher risk (especially in the period 15-40 months from time zero) and the rest switchings to the low risk first latent state (possibly due to successful medical treatment). Finally, a third group of patients is at very high risk of high NYHA class at baseline time. Most of them remain at high risk during follow-up, but a slight proportion switches to better propensity states; surprisingly more often to state 1 than to state 2. This might be due to increased medical attention given to high risk patients.

References

1. Bartolucci, F., Farcomeni, A.: A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. Biometrics **71**, 80–89 (2015)
2. Bartolucci, F., Farcomeni, A., Pennoni, F.: Latent Markov Models for Longitudinal Data. Chapman & Hall/CRC Press, Boca Raton, FL (2013)
3. Böckenholt, U.: A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. Psychological Methods **10**, 65–83 (2005)

Fig. 1 *MDCM data: estimated transition matrix as a function of time t when* $k = 3$



4. Gigli, M., Stolfo, D., Merlo, M., Barbati, G., Ramani, F., Brun, F., Pinamonti, B., Sinagra, G.: Insights into mildly dilated cardiomyopathy: Temporal evolution and long-term prognosis. European Journal of Heart Failure **19**, 531–539 (2017)
5. Liggett, T.M.: Continous Time Markov Processes. American Mathematical Society, Providence (2010)
6. Rizopoulos, D., Verbeke, G., Molenberghs, G.: Shared parameter models under random effects misspecification. Biometrika **95**, 63–74 (2008)