Lexis Random Fields Modeling in Mortality Surface Analysis

Fabio Divino[1]

[1] Division of Physics, Computer Science and Mathematics,
University of Molise, Italy

ABSTRACT Mortality surfaces are intensity measures of mortality at population level as function of age and time, they represent an important statistical tool to explain and forecast the dynamics of mortality. Since the paper by Vaupel et al. (1979) different approaches have been proposed, see for example Lee and Carter (1992). The main goal is to take into account the joint effects of period, age and cohort and their possible interactions.

In this work we develop an approach based on Markov random fields to estimate the mortality intensity and to decompose it into two useful interpretable components. The use of Markov random fields in spatial statistics originated in the framework of *image analysis* during the 80s and in the last 20 years has become very popular in many fields of application as in ecological analysis, disease mapping and biogeography (see Besag et al., 1991).

In simple words, a random field is a joint probability model for a set of n random variables positioned on the knots of a finite graph, with a given dependence structure. When the dependence is Markovian the random field is a Markov random field (MRF). Let $\mathcal{T}$ domain of the time variable t and $\mathcal{X}$ the domain of the age variable x, in this framework the random quantities of interest are represented by the mortality intensities $\mu_{t,x}$ related to the cells (t,x) of a 2-dimensional *Lexis* diagram $\mathcal{L} = \mathcal{T} \times \mathcal{X}$ (Keiding, 1991).

Our approach consists of two levels. The first one describes the hidden status variables underlying the data. At this level we model the mortality intensities $\mu_{t,x}$ through a random field with two separable components. At the second level, the conditional likelihood model describes the observable death counts $D_{t,x}$, considered as random variables independently Poisson distributed with mean $\mu_{t,x} N_{t,x}$, where $N_{t,x}$ is the size of the population at risk concerning the cell (t,x).

In general, we expect that the mortality intensity is rather smooth across the *Lexis* diagram. Thus, we assume that the quantities $\mu_{t,x}$ should not vary much in time and age and present a smooth pattern. This aspect is modelled through the use of a random field with Markovian dependence. On top of a main smooth pattern, we should account also for additional mortality in excess or in defect. This second feature, that could be the result of latent factors, can be modelled by a second component of independent random effects. Under these assumptions, for each cell $(t,x) \in \mathcal{L}$, we define the following probabilistic model for $\mu_{t,x}$

$$\log \mu_{t,x} = \rho_{t,x} + \zeta_{t,x}. \tag{1}$$

In particular, we assume that the joint distribution of $\rho = \{\rho_{t,x} : (t,x) \in \mathcal{L}\}$ is a Gaussian random field with interaction parameter γ and Markovian neighbourhood system Δ (GMRF). More precisely, the joint density of ρ is given, less than a normalizing constant, by $\pi(\rho) \propto \exp\{-\gamma \sum_{(t,x)\sim(t',x')}(\rho_{t,x} - \rho_{t',x'})^2\}$. This is the classical pairwise smoothing Gaussian Markov random field which aims similarity of $\rho_{t,x}$ with other variables $\rho_{t',x'}$ located in its neighbourhood, denoted by the cells $(t',x') \sim (t,x)$. The random effects $\zeta_{t,x}$ are assumed to be independent *Gaussian* distributed with mean and precision equal respectively to 0 and σ^{-1}.

In the model (1) $\rho_{t,x}$ follows a *Lexis* smooth model while the second component $\zeta_{t,x}$ is the additional random effect necessary to recompose the full mortality intensity $\mu_{t,x}$. More precisely, the GMRF models the *Lexis* dynamic affecting the mortality, given by the smooth joint effect of period and age, and also induces smoothness between neighbourly cohorts. The second component, the independent random effects, accounts for the possible effect of latent factors. Typical effects included in that component are the well known cohort frailty effects.

We approach the inferential problem in a Bayesian setting: the hidden model represents our *prior* guess on the quantities of interest $\mu_{t,x}$, while the *likelihood* includes the empirical evidence of the data. The parameters can be estimated through the *posterior* model resulting by joining *prior* with *likelihood* through the Bayes rule. The marginal profiles of the full posterior distribution can be simulated by Markov chain Monte Carlo techniques to compute estimates and credibility intervals.

Applications on real data have shown fitting and flexibility of the proposed approach as well as the possibility to interpret the mortality in two separable components: the smoothing by period-age-cohort and the heterogeneity by frailty and latent factors.

Keywords: Hidden model, Lexis diagram, Markov random field, MCMC, mortality surface .

References:

Besag, J., J.C. York, A. Mollie (1991). Bayesian image restoration, with applications in spatial statistics. *Annals of the Institute of Statistical Mathematics. Vol. 43, n. 1 pp. 1-59.*

Keiding N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transaction of the Royal Society A. Vol. 332, n. 1627, pp. 487-509.*

Lee, R.D., L.R. Carter (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association Vol.87, n. 49, pp. 659-671.*

Vaupel, J.W., K.G. Manton, E. Stallard (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography. Vol. 16, n. 3, pp. 439-454.*