Modelling mortality data on the Lexis diagram

James Kirkby and Iain Currie

1 Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland
Email: jgk1@hw.ac.uk

Abstract: Currie, Durban & Eilers (2004) used 2-dimensional \( P \)-splines to smooth mortality data classified by age at death and year of death. In this paper we apply this model to data classified by age at death, year of death and year of birth. Discrete cohort effects are added to the model using a method similar to the overdispersion model of Perperoglou and Eilers (2006). This model allows us to decompose the mortality surface into (i) a smooth 2-dimensional surface in age and time and (ii) discrete cohort effects. We illustrate our remarks with the analysis of some German mortality data.

Keywords: Lexis diagram; Mortality; Overdispersion; \( P \)-splines.

1 Description of the data

We suppose we have census data on 1st January for years \( t = 1, \ldots, n_y \) which give the number of lives with ages between \( x \) and \( x + 1 \) for ages labelled \( x = 1, \ldots, n_a \); these data are held in \( E = (e_{x,t}), n_a \times n_y \). Let \( d_{x,t} \) be the number of deaths from the \( e_{x,t} \) lives in the calendar year \( t \) to \( t + 1 \). These deaths can occur between ages \( x \) and \( x + 1 \) (triangles of type \( A \) in Fig. 1), or between ages \( x + 1 \) and \( x + 2 \) (triangles of type \( B \)); triangles \( A \) and \( B \) are known as Lexis triangles (Carstensen & Keiding, 2005). Let \( D_A = (d_{A,x,t}), n_a \times n_y \), hold the number of deaths in triangles of type \( A \) and \( D_B = (d_{B,x,t}), n_a \times n_y \), hold the number of deaths in triangles of type \( B \). We assume that the three matrices \( E, D_A \) and \( D_B \) are available. We seek estimates, \( E_A \) and \( E_B \), of the total times lived in triangles \( A \) and \( B \) respectively. Under the assumption that the \( d_{A,x,t} \) and \( d_{B,x,t} \) deaths are distributed uniformly over \( A \) and \( B \) respectively it is then straightforward to show that the expected times lived are

\[
E_A = \frac{1}{2} E - \frac{1}{6} D_A - \frac{1}{6} D_B, \quad E_B = \frac{1}{2} E - \frac{1}{4} D_A - \frac{1}{4} D_B. \quad (1)
\]

For a life that dies in \( A \) the expected time of death is \( t + \frac{1}{3} \) and the expected age at death is \( x + \frac{2}{3} \); similarly, for a life that dies in \( B \) the expected time of death is \( t + \frac{1}{4} \) and the expected age at death is \( x + \frac{4}{3} \). The situation is summarised in Fig. 1. We now model the number of deaths as follows

\[
d_{A,x,t} \sim \mathcal{P}(e_{A,x,t} | \mu_{x+2/3,t+1/3}) \quad (2)
\]
FIGURE 1. Lexis diagram: $E$ lives between age $x$ and $x + 1$ at year $t$, $D_A$ deaths and $E_A$ years lived in triangle $A$, $D_B$ deaths and $E_B$ years lived in triangle $B$.

\[ d_{B,x,t} \sim \mathcal{P}(e_{B,x,t} \mu_{x+4/3,t+2/3}) \] (3)

where $\mu_{x,t}$ is the hazard rate at age $x$ and time $t$. In this model the deaths and exposures in triangles of type $A$ are located on a rectangular grid. The same remark applies to the data in triangles of type $B$ so the whole data set consists of two interleaved grids. We seek simple forms for the function $\mu_{x,t}$ which decompose the mortality surface into (a) a smooth 2-dimensional surface and (b) discrete cohort and period effects. The smooth 2-dimensional surface represents underlying mortality while the discrete effects represent shocks to the surface caused, for example, by a cold winter or a flu epidemic (period effects) or by some effect at birth which influences the future mortality of that cohort. It is of particular interest to identify any such cohort effects.
2 A 2-dimensional smooth surface

We use the method of P-splines (Eilers & Marx, 1996). We take two rich bases of equally spaced cubic $B$-splines, one to cover the $x$-values from \(1 + \frac{2}{3}\) to \(n_a + \frac{1}{3}\) and another to cover the $t$-values from \(1 + \frac{1}{3}\) to \(n_y + \frac{2}{3}\). These bases give rise to marginal regression matrices $B_{A_a}$ and $B_{A_y}$ for the ages and years in triangles of type $A$. Since these data points lie on a grid, $B_A = B_{A_a} \otimes B_{A_y}$ is the regression matrix for a 2-dimensional surface; together with (2) we have defined a generalized linear model for data in triangles of type $A$. Similarly, $B_B = B_{B_a} \otimes B_{B_y}$ is a regression matrix for the data in triangles of type $B$. Finally $B = [B_A : B_B]$ is the regression matrix of a surface for the complete data set. Smoothness of the fitted surface is ensured by marginal penalization and the model is fitted by penalized likelihood.

Fig. 2 shows a plot of the observed mortality surface (on the log scale) for some German mortality data (Statistisches Bundesamt). The obvious diagonal crests and troughs in Fig. 2 are indications of cohort effects. Fig. 2 also shows the surface fitted by the model described in the previous paragraph. The diagonal crests in the raw data are visible in the smooth surface, but
Fig. 2 suggests an alternative model where the individual cohorts effects are modelled as separate additive effects rather than as part of the underlying smooth mortality surface.

3 Adding cohort effects

We extend the basic smooth model to include cohort effects, as follows

$$\eta = Ba + C\gamma$$

(4)

where $C$ is the design matrix for the individual cohort effects. A ridge penalty with smoothing parameter $\kappa$ is applied to $\gamma$. The ridge penalty maintains identifiability and the size of $\kappa$ can be tuned to the observed cohort effects. The ridge penalty ensures that the smooth features of the data are described by the $B$-spline surface and only the additional variation caused by the cohort effects is absorbed by $\gamma$.

For given values of the smoothing parameters, estimates of $a$ and $\gamma$ are obtained by solving:

$$
\begin{bmatrix}
B'WB + P_1 & B'WC \\
C'WB & C'WC + P_2
\end{bmatrix}
\begin{bmatrix}
a \\
\gamma
\end{bmatrix}
= 
\begin{bmatrix}
B'\tilde{W}\tilde{z} \\
C'\tilde{W}\tilde{z}
\end{bmatrix}
$$

(5)

where $\tilde{z}$ is the usual working vector, and $P_1$ and $P_2$ are the appropriate difference and ridge penalty matrices. The values of the smoothing parameters are chosen by BIC.

Triangles $A$ and $B$ in the same cohort in Fig. 1 are modelled by a single parameter in model (4). However, lives in triangle $A$ are born predominantly in the first half of the year while those in triangle $B$ are born predominantly in the second half. Thus we can split the cohort effect into two parts. We
extend model (4) and use separate parameters, one for each of the two types of triangle within a year of birth; the linear predictor becomes

$$\eta = Ba + C_A \gamma_A + C_B \gamma_B.$$  \hspace{1cm} (6)

This model gives a much improved fit (as measured by the BIC value). Figs. 3 and 4 show that this model successfully decomposes the mortality surface into a smooth 2-dimensional surface and a set of cohort effects. Period effects can be added in the same way as (4) but were not found to be significant for this data set.

There are some connections between our cohort models (4) and (6) and the overdispersion models of Perperoglou & Eilers (2006). These authors account for overdispersion by fitting an extra parameter for each data point. Their approach differs from the usual method of over-dispersion modelling which assumes the specification of the variance is not correct. In the standard approach estimation usually proceeds by using quasi-likelihood to re-specify the mean-variance relationship. The two approaches correspond to different underlying structures in the data: the Perperoglou & Eilers approach would be appropriate in the case of an underlying smooth trend shocked by individual random effects, while the second corresponds to a genuinely over-dispersed distribution for each data point. In one dimensional problems, it is difficult to distinguish between the two structures, and the quasi-likelihood approach has the advantage of being able to deal with under-dispersion. In higher dimensional problems, as here, we can distinguish between the two structures provided that the shocks follow some...
kind of systematic pattern. In our case, the cohort patterns were clearly visible in the data, and the basic smooth model was not satisfactory, so an addition to the linear predictor was natural. In other situations close inspection of the data may lead to identification of additive effects that may otherwise be put down to over-dispersion.

It is also possible to fit model (4) by expressing it as mixed model. Re-parameterizing the $P$-spline model using the method given by Currie et al. (2006), the similarity between the system in (5) and the Fisher scoring algorithm given by Breslow & Clayton (1993) becomes clear. The smoothing parameters could then be optimized using ML or REML.

4 Conclusions

The richness of data on the Lexis Diagram allowed us to examine cohort effects in detail. For the data in our example we found strong evidence of cohort effects of two kinds: cohort effects, $\gamma_A$, associated with triangles of type A (early year births) and distinct cohort effects, $\gamma_B$, associated with triangles of type B (late year births).

The data set used in this paper was relatively small (12 years and 80 ages), and we were able to fit the model using standard matrix computations. If, for a bigger data set, the computations became unmanageable, an efficient algorithm to solve systems similar to (5) can be found in Perperoglou & Eilers (2006). Further gains are possible by taking advantage of the Kronecker product structure of the $B$-spline basis (Currie et al., 2006).

References


Statistisches Bundesamt, Various Tables (Data obtained through the Human Mortality Database, www.mortality.org, December 2005.)