Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics and Data Analysis 52 (2008) 4346–4356

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

A bivariate frailty model for events with a permanent survivor fraction and non-monotonic hazards; with an application to age at first maternity

Peter Congdon*

Department of Geography, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

Received 30 July 2007; received in revised form 13 February 2008; accepted 13 February 2008
Available online 10 March 2008

Abstract

For certain life cycle events a non-susceptible fraction of subjects will never undergo the event. In demographic applications, examples are provided by marriage and age at first maternity. A model for survival data allowing a permanent survival fraction, non-monotonic failure rates and unobserved frailty is considered here. Regressions are used to explain both the failure time and permanent survival mechanisms and additive correlated errors are included in the general linear models defining these regressions. A hierarchical Bayesian approach is adopted with likelihood conditional on the random frailty effects and a second stage prior defining the bivariate density of those effects. The gain in model fit, and potential effects on inference, from adding frailty is demonstrated in a case study application to age at first maternity in Germany.

1. Introduction

The majority of survival analysis applications assume a survivor function that tends to zero as the timescale governing the occurrence of events increases. However, in some types of life cycle event there will be a fraction of subjects who will never undergo the event, or in a terminology sometimes used, are non-susceptible (Hougaard et al., 1994). In demographic applications, examples are provided by marriage and age at first maternity. Not allowing for permanent survival when it exists may distort estimates of important summary parameters such as modal time of event and maximum event rate. A further distinguishing feature of many demographic and biological processes is that the event rate is non-monotonic, for example, the rate at first increases but after reaching a peak tails off again (Gore et al., 1984). Parametric models accommodating such a pattern include the log-logistic and Burr XII densities, and sickle models (Bennett, 1983; Diekmann and Mitter, 1983).

This paper considers a model for survival data with a permanent survival fraction and non-monotonic failure rates and evaluates the gain in model fit, and effects on inference, from adding frailty. An application considers age at first maternity using data from the 2002 German General Social Survey, with permanent survival amounting to childlessness. Regressions are used to explain both the failure time of the event (here age at first maternity) and the

* Tel.: +44 20 7 882 5400; fax: +44 20 8981 6276.
E-mail address: p.congdon@qmul.ac.uk.

0167-9473/$ - see front matter
doi:10.1016/j.csda.2008.02.017

permanent survival mechanism (susceptibility to undergo maternity or not). Additive correlated effects are included in the linked models defining these regressions and relate to two types of frailty: influences on the event rate itself and influences on the probability of susceptibility.

A hierarchical Bayesian approach is adopted with likelihood conditional on bivariate random frailty effects, and a second stage prior defining the density of those effects. A Bayesian approach facilitates modelling with multivariate random effects whereas frequentist approaches based on marginal likelihoods – with random effects integrated out using numerical methods – may become infeasible or unreliable when there are many random parameters (Tutz and Kauermann, 2003; Kim et al., 2002). Monte Carlo Markov Chain (MCMC) methods are used for estimation via the WINBUGS package (Lunn et al., 2000), and generate samples from the posterior distribution without the form of the posterior density being known analytically (Gilks et al., 1996). This is useful in summarizing possibly non-normal densities relating to model functionals (e.g. modal ages at maternity) and in obtaining posterior probabilities relating to hypotheses on such functionals, e.g. that the modal age for women with low education years is lower than the modal age for women with extended education. This benefit is considered more fully in Section 4.

2. Specifications for permanent survival and non-monotonic hazards

The usual approach to modelling events with a permanent survival fraction (the binary PSF) assumes the total survival rate is a binary mixture in which the subpopulation of permanent survivors has survivor rate $S_p(t) = 1$, and the other (the susceptible subpopulation subject to depletion) follows conventional survival with $S_n(t) \rightarrow 0$ as $t \rightarrow \infty$. The overall survivor function is

$$S^*(t) = (1 - \pi)S_p(t) + \pi S_n(t) = (1 - \pi) + \pi S_n(t)$$

where π is the probability of susceptibility or propensity to experience the event. Let R_i be a partially unobserved binary indicator with $R_i = 1$ if a subject is susceptible, with $Pr(R_i = 1) = \pi_i$. Denote the density, survival and hazard functions for susceptibles simply as $f(t)$, $S(t)$ and $h(t)$. Then for subjects observed to fail (denoted by the binary indicator $d_i = 1$), it follows that $R_i = 1$, with likelihood contribution

$$Pr(R_i = 1) f(t_i) = \pi_i f(t_i).$$

Censored subjects (with $d_i = 0$) may be either susceptibles or non-susceptibles with likelihood contribution

$$Pr(R_i = 0) + Pr(R_i = 1)Pr(T > t_i) = (1 - \pi_i) + \pi_i S(t_i).$$

The total likelihood contribution is then

$$[\pi_i f(t_i)]^{d_i}[(1 - \pi_i) + \pi_i S(t_i)]^{1-d_i},$$

with the hazard for the combined population being

$$\frac{\pi_i f(t_i)}{(1 - \pi_i) + \pi_i S(t_i)} = \frac{\pi_i h(t_i)S(t_i)}{(1 - \pi_i) + \pi_i S(t_i)}.$$

Chen et al. (1999) describe an alternative scheme (here called the Poisson PSF) in which for subject i there is a latent Poisson count of risks R_i, with mean θ_i, and with unobserved times $U_{i1}, \ldots, U_{iR_i}$ associated with each of these risks. The U_{ir} follow the same failure distribution $F(t) = 1 - S(t)$, and the observed failure time t_i is the minimum of these times. If $R_i = 0$ then a subject survives permanently from the event being modelled with probability $\exp(-\theta_i)$.

For an event with non-monotonic hazard (e.g. age at first maternity) the log-logistic failure distribution is appropriate as it allows an initial increase and subsequent decrease in chances of the event (Bennett, 1983). This distribution has hazard and survivor functions

$$h(t) = \frac{\alpha\lambda^\alpha t^{\alpha-1}}{[1 + (\lambda t)^\alpha]},$$
$$S(t) = [1 + (\lambda t)^\alpha]^{-1},$$

and corresponding density

$$f(t) = \frac{\alpha\lambda^\alpha t^{\alpha-1}}{[1 + (\lambda t)^\alpha]^2}$$