

Modelling Human Mortality Trajectories via an Accelerated-Ageing Hazard Function¹

Un modello di fragilità a rischi non proporzionali per l'analisi della sopravvivenza umana

Elisabetta Barbi

Dip. Scienze Demografiche

La Sapienza

elisabetta.barbi@uniroma1.it

Francesco Lagona

Dip. Istituzioni Politiche e Scienze Sociali

Roma Tre

lagona@uniroma3.it

Riassunto: Si propone un *frailty model* a rischi non proporzionali per l'analisi della sopravvivenza della popolazione umana, ottenuto accelerando la funzione di rischio individuale con una seconda componente di fragilità non osservata. Rispetto al *frailty model* a rischi proporzionali, le stime prodotte suggeriscono una diversa spiegazione della decelerazione osservata nei rischi di mortalità a livello di popolazione e possono condurre ad una diversa valutazione dei limiti della longevità umana. La stima di massima verosimiglianza del modello è realizzata mediante un algoritmo E-M con passo E valutato mediante *importance sampling*.

Keywords: accelerated-ageing hazard, frailty models, importance sampling, E-M algorithm.

1. Introduction

It has been acknowledged that fitting *population* mortality data by using a model relating to *individual* hazards may well lead to serious misunderstandings of the results, if individuals do not share the same risk of mortality. For example, a popular model for the mortality hazard rate or force of mortality $\mu(x)$ - x denoting lifetime - is the Gompertz function: it is supported by biological theories of individual ageing, but fits poorly mortality data at old ages, because observed hazards decelerate around age 80, following a logistic pattern. Vaupel *et al.* (1979) explain that cohort hazards decelerate because individuals differ in their vulnerability (frailty) to death; hence, the selection process towards low-frailty individuals changes the composition of the cohort during its life and, as a result, the (cohort) hazard trajectory decelerates at old ages. In technical terms, the (observed) mortality hazard rate $\bar{\mu}(x)$ might be viewed as a mixture of (unobserved) individual mortality hazards, where the mixing weights relate to the distribution of the frailty in the cohort. Formally, they (1) modelled the individual force of mortality as a relative-risks model $\mu(x; z) = z\mu(x)$, where $\mu(x)$ is the baseline hazard and (2) proved that, if z is Gamma distributed with unit expectation, then $\bar{\mu}(x) = \mu(x) [1 + \sigma^2 H(x)]^{-1}$ where $\sigma^2 = \text{Var}z$ and $H(x)$ is the baseline cumulative hazard. Accordingly, ecological inferences on the basis of population mortality are possible only when $\sigma^2 = 0$, i.e. when individ-

¹Support for the present work is gratefully acknowledged. Barbi is supported by the Young Researcher Program 2000 (University La Sapienza of Rome) "Specification and estimation of statistical models for human survival" and Lagona is supported by the MURST Grant 2002 (University Roma Tre) "Frailty models for the statistical analysis of mortality data at old ages".

ual mortality risks are homogeneous. Furthermore, if a Gompertz function is used as a baseline, the mean force of mortality follows a logistic shape, as observed.

More realistically, it should be taken into account that the frailest individuals in the population may suffer an increasing debilitation with age (Wilmoth and Horiuchi, 1999). Hence, beside the heterogeneity captured by the proportional hazard hypothesis, a second source of heterogeneity should be addressed, due to possible individual differences in the rate of ageing. This hypothesis has motivated the acceleration of the classic frailty model by entering a second frailty component that operates multiplicatively on the lifetime.

Unfortunately, a random term acting multiplicatively on age makes nor the predictive distribution of frailties neither the probabilities of death at the population level available in closed form. Although demographers agree on the limit of the relative-risks hypothesis in the analysis of human survival, this technical issue has discouraged the wide spread of studies involving the non-proportional risks assumption. To overcome the problem, we suggest and implement a Monte Carlo E-M algorithm.

2. An accelerated-ageing model

Let X denote the random lifetime x of an individual belonging to a human population. Given two random effects, or frailties, (z_1, z_2) , we assume that conditional on frailties lifetimes are independent and the force of mortality of each individual takes the form

$$\mu(x; z_1, z_2) = z_1 \alpha \exp(z_2 \beta x) + \gamma \quad (1)$$

where the baseline hazard

$$\mu(x; 1, 1) = \alpha \exp(\beta x) + \gamma \quad (2)$$

is the Gompertz-Makeham hazard function. Additionally, we assume that z_1 and z_2 are independently sampled from two Gamma densities $g_k(z_k)$, $k = 1, 2$, with $EZ_k = 1$ and $\text{Var}Z_k = \sigma_k^2$ (identifiability condition; Vaupel *et al*, 1979). Under (1), the cohort force of mortality takes the form

$$\bar{\mu} = \int \bar{\mu}(z_2) g_2(z_2) dz_2 + \gamma \quad (3)$$

where

$$\bar{\mu}(z_2) = \frac{\alpha \exp(z_2 \beta x)}{1 + \sigma_1^2 H(x, z_2)} + \gamma$$

and

$$H(x, z_2) = \int_0^x \mu(t; 1, z_2) dt$$

are the mean force of mortality and the cumulative hazard of the cohort stratum where individuals share $Z_2 = z_2$, respectively. Given the ML estimates of the parameters $\theta = (\alpha, \beta, \gamma)$ and $\sigma = (\sigma_1^2, \sigma_2^2)$, (3) can be evaluated by drawing samples from $g_2(z_2)$.

Given the observation of d_x deaths among l_x survivors at age x , for ages, say, $x_0, x_0 + 1, \dots, \omega$, the marginal loglikelihood is given by

$$\sum_{x=x_0}^{\omega} \{d_x \log \bar{q}(x) + (l_x - d_x) \log [1 - \bar{q}(x)]\}$$

where

$$\bar{q}(x) = 1 - \exp \left\{ - \int_x^{x+1} \bar{\mu}(t) dt \right\}$$

is the probability of death for an individual of age x . The intractability of $\bar{q}(x)$ has motivated the use of an E-M algorithm with a Monte Carlo E step. Following the missing data terminology, assume to observe the z_2 -frailty values collected in a vector $\mathbf{z}_x$, $x = x_0, x_0 + 1, \dots, \omega$. At the n th step, the complete data loglikelihood takes the form

$$l(\boldsymbol{\theta}, \boldsymbol{\sigma}) = \sum_{x=x_0}^{\omega} (d_x \log [\bar{q}(\mathbf{z}_x)] + (l_x - d_x) \log [1 - \bar{q}(\mathbf{z}_x)]) + \sum_{x=x_0}^{\omega} l_x \log g_x(\mathbf{z}_x) \quad (4)$$

where $g_x(\mathbf{z}_x)$ is the predictive distribution of the z_2 -frailty values among the l_x survivors. To evaluate the expectation (E step)

$$Q\left((\boldsymbol{\theta}, \boldsymbol{\sigma}), (\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\sigma}^{(n-1)})\right) = \int l(\boldsymbol{\theta}, \boldsymbol{\sigma}) g(\mathbf{z}, \boldsymbol{\theta}^{(n-1)}, \boldsymbol{\sigma}^{(n-1)}) d\mathbf{z}$$

importance sampling can be used by drawing samples from the z_2 -frailty predictive distribution computed at the point estimates $(\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\sigma}^{(n-1)})$ given by the previous $(n-1)$ th M step. If enough samples are drawn, then these Monte Carlo E-M iterations converge to the maximum of the loglikelihood (Booth and Hobert, 1999). In particular, we suggest to automatically increase the number of samples when approaching the target, thus gaining absolute convergence of the algorithm. Finally, the Louis' (1982) observed information matrix can be obtained as a side product from the samples in the last iteration and asymptotic standard errors of the MLEs can be easily computed.

3. Application

To give an example, the accelerated-ageing and the relative-risks models have been estimated on a Swedish data set. The models are fitted to probabilities of death for women belonging to the extinct cohorts born between 1861 and 1870, from age 60 to age 95 (data from the Berkeley Mortality Database, <http://demog.berkeley.edu/wilmoth/mortality>).

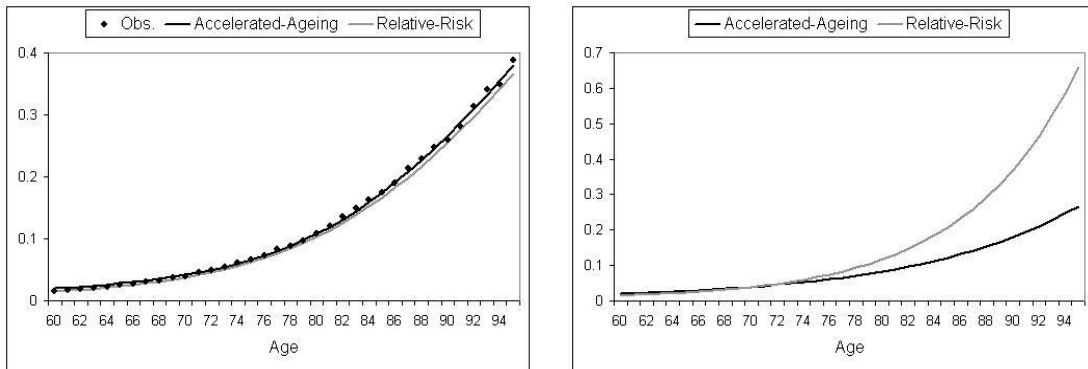


Figure 1 : left: estimated and observed $\mu(x, 1, 1)$; right: estimated $\mu(x, 1, 1)$. Swedish female cohorts 1861-70.

By choosing the starting age 60 we implicitly assume that frailty is fixed at this age. As a consequence, the estimated variance in frailty arises from genetic attributes, early-life events and other possible environmental factors acquired by the individuals up to age 60. The age 95 was chosen as the upper limit for the fitting process because of the many fluctuations of data over this age.

Figure 1 depicts the results of the fitting process for both the models: figure 1(left) shows the estimated mortality curves at the population level whereas figure 1(right) shows the relative curves at the individual level. Although both the models seem reasonable and give significant estimates (see Table 1), the accelerated-ageing model seems to give a better fit compared to that given by the relative-risks model. The estimated parameters support the conclusion that individual differences increasing with age may play an important role in determining the mortality trajectory of population at old ages. Furthermore, the hazard trajectory of the standard individual (Figure 1, right) rises lower than that estimated by the relative-risks model. As a consequence, the mortality curves for robust individuals estimated through the accelerated-ageing risks are more favorable to those estimated by the relative risks model. In other words, the nonproportional risks model considered here gives results that might match better with the observation of extreme ages at death (e.g., that of Jeanne Calment, the famous French woman died in 1997 at age 122.5) that have never been captured by other mortality models neither by the proportional risks frailty model (Barbi *et al.*, 1999).

Table 1: MLEs of the relative and accelerated-ageing models. Swedish female cohorts 1861-70.

		α	β	γ	σ_1^2	σ_2^2
<i>relative risks</i>	mle	.0109	.1169	.0041	.1444	-
	se	.0004	.0019	.0005	.0115	-
<i>accelerated risks</i>	mle	.0160	.0801	.0043	.0622	.1989
	se	.0002	.0014	.0002	.0011	.0025

References

- Barbi E., Caselli G. and Vallin J. (1999) Trajectories of extreme survival in heterogeneous population, *PAA Annual Meeting, New York*.
- Booth J.G. and Hobert J.P. (1999) Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm, *Journal of the Royal Statistical Society, B*, 61, 265–285.
- Louis T. (1982) Finding observed information using the em algorithm, *Journal of the Royal Statistical society, B*, 44, 98–130.
- Vaupel J.W., Manton K.G. and Stallard E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, 16, 439–454.
- Wilmoth J. and Horiuchi S. (1999) Do the oldest old grow old more slowly?, in: *The Paradoxes of Longevity*, B. Forette C. Franceschi J.M.R. and Allard M., eds., Springer Verlag, Heidelberg, 35–60.