A mortality model based on a mixture distribution function

Un modello di mortalità basato su una mistura di distribuzioni

Stefano Mazzuco and Bruno Scarpa and Lucia Zanotto

Abstract A new mortality model based on a mixture distribution function is proposed. We mix a half-normal distribution with a generalization of the skew-normal distribution. As a result we get a six-paramters distribution function having a good fit with a wide variety of mortality patterns. This model is fitted to several mortality data and compared with the Siler model (five parameters). The main feature of the proposed model is that it has a relatively good fit even in mortality pattern with a high "accident" hump (this happens, for instance, in countries undergoing armed conflicts or with a high HIV prevalence), while the Siler model cannot catch this hump.

Abstract *Si propone un nuovo modello di mortalità basato sun una mistura tra una distribuzione half-normal e una generalizzazione della skew-normal. Come risultato otteniamo una funzione di densità a sei parametri che mostra un buon adattamento con svariati pattern di mortalità. Il modello viene confrontato con il modello di Siler (5 parametri) su vari dati di mortalità. Il maggior pregio del modello proposto è il suo adattamento relativamente buono anche nei casi in cui l'incidenza della mortalità accidentale (ad esempio, nei paesi in guerra o con un'alta incidenza di HIV) alta, mentre il modello di Siler non riesce a cogliere questo aspetto.*

Key words: Mortality model, mixture distribution , skew-normal distribution

---

Stefano Mazzuco
Dipartimento di Scienze Statistiche, Via Cesare Battisti, 241, 35121 Padova, e-mail: mazzuco@stat.unipd.it

Bruno Scarpa
Dipartimento di Scienze Statistiche, Via Cesare Battisti, 241, 35121 Padova, e-mail: scarpa@stat.unipd.it

Lucia Zanotto
Dipartimento di Scienze Statistiche, Via Cesare Battisti, 241, 35121 Padova, e-mail: zanotto@stat.unipd.it

1 Introduction

Modeling mortality data is challenging as there are many features of mortality patterns that are difficult to incorporate into a single model. At the risk of over simplifying we can say that we have on the one hand the Gompertz model (and its extension by Makeham) that is compact, with few parameters (only 3) and relatively easy to estimate and on the other hand the Heligman-Pollard model having 8 parameters, which makes it able to catch all the characteristics of mortality pattern but also very difficult to estimate due to identifiability problems posed by the high number of parameters. In between these two solutions the Siler model [6, 7] – which is a generalization of the Gompertz-Makeham model – has 5 parameters and a relatively good fit with several mortality patterns, although what is generally referred to as "accident" hump (i. e. a hump often observed among males deaths function in correspondence of young adulthood ages) cannot be caught. We therefore propose a new model, based on a mixture distribution function, having six parameters with the possibility of incorporating the "accident" hump.

2 Modelling the "deaths" function

The underlying idea of our proposal is that deaths rather than rates (as in the Gompertz-Makeham model) or probabilities (as in the Heligman-Pollard model) are modelled. This choice stems from the fact that deaths of a life table can be seen as a probability density function. We therefore need to find a density function that can have a good fit with deaths of a life table. This means finding a probability density function with at least two modes (one at birth and one at the Lexis point) but with possibly a third one, catching the so-called "accident hump". This is said to occur because pubescent and young adult boys suddenly become much more likely to die in accidents and violence. However, this hump can also be observed among women, for instance in countries with a particularly high prevalence of HIV epidemic. The Heligman-Pollard model [3] takes into account this feature but is extremely difficult to fit. Therefore, we define a mixture between a half-normal variable and a generalization of a skew-normal variable.
The half normal variable has the following distribution function

$$f_Y(y;\sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right) \quad y > 0 \tag{1}$$

which has a maximum in zero, and then decreases as y increases.
Rocha *et al.* [4] propose a generalization of the skew-normal distribution, which characteristics are discussed by Elal-Olivero *et al.* [2]. Its probability density function is the following

$$f_X(x;\xi,\omega^2,\alpha,\lambda) = 2\omega^{-1}\left(\frac{1+\alpha\left(\frac{x-\xi}{\omega}\right)^2}{1+\alpha}\right)\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left\{\lambda\left(\frac{x-\xi}{\omega}\right)\right\}. \quad (2)$$

where $b_2 = \xi^2 + \omega^2$. The (2) can have - at most - two modes depending on the value of α. In particular if $\alpha \geq 0$, the resulting pdf has two modes, otherwise ($\alpha < 0$) the pdf is unimodal.
The half-normal distribution is a good candidate to catch mortality pattern at early ages, while the skew-bimodal-normal (SBN) distribution can catch mortality at adult ages, and if an accident hump exists this can be catched by the second mode of the SBN distribution. In order to make a mixture distribution out of these two, we need to truncate at 0 the FGSN pdf. So the mixture distribution function is defined as

$$f_Z(z;m,\sigma,\xi,\omega^2,\alpha,\lambda) = m \cdot f_Y(z;\sigma) + (1-m)\cdot f_X(z;\xi,\omega^2,\alpha,\lambda) \quad (3)$$

Although the mixture distribution function might be look awkward, we find a sensible demographic interpretation for all the parameters:

- m is the mixture parameter, telling us the relative weight of early ages mortality. If m is high, infant mortality is high.
- σ is the variance of the half-normal distribution, telling us to what extent the peak of mortality at birth is extended to ages greater than 0. We can say that σ and m indicate us the degree of "horizontalization" of the survival curve (see Cheung et al [1])
- ξ is the location parameter of the FGSN distribution, and it tells us something about the longevity of the adult population.
- ω is the scale parameter of the FGSN distribution, telling us how much adult mortality is concentrated in the Lexis point. We can say that ω indicates the degree of "verticalization" of the survival curve (see Cheung et al [1]).
- α is a shape parameter of the SBN distribution. In particular, it is related with the distribution unimodality. If $\alpha \geq 0.5$ then the pdf is bimodal, while if $0 < \alpha < 0.5$ the distribution is unimodal.
- β is a shape parameter on which value depends the skewness of the distribution. In particular, if λ is positive then the pdf is skewed on the right, while if λ is negative the distribution is skewed on the left.

2.1 Comparison with Siler model

We compare model (3) with the Siler [6, 7] model. This is a five parameter model – basically, an extension of the Gompertz-Makeham model, in which the hazard function is as follow

$$h(t;\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2) = \alpha_1 e^{-\beta_1 t} + \alpha_2 + \alpha_3 e^{\beta_2 t}. \quad (4)$$

(4) can be seen as the sum of three independent hazard function being the first one ($\alpha_1 e^{-\beta_1 t}$) the hazard of "immature death" and the second and third ones the Gompertz-Makeham hazard function describing adult mortality. This means that α_1 and β_1 are related with infant mortality while α_2, α_3 and β_2 are related with adult and old age mortality. Therefore there is no parameter taking into account a possible accident hump. In contrast, the Heligman-Pollard model has three parameters related to the accident hump, but its high number of paramters makes it very difficult to estimate it unless a Bayesian approach is considered as in Sharrow et al [5].

3 Preliminary results

We fit the mixture model (3) and the Siler model described in (4) to some real data. We consider mortality data from Human Mortality Database [8] and from Who mortality database. In this abstract we show Figures 1 and 2 describing models fit with USA 2002 (men) and Botswana 2011 (women) mortality pattern. They both have a accident hump even though in the case of Botswana the hump is not due to accidents but to HIV incidence among women. We can see that in both cases Siler model cannot catch these humps while the mixture model does. In the next future, the mixture model will be compared with Heligman-Pollard model, too. Our expectation is that Heligman-Pollard model provides the best fit but not too far from what we get from the mixture model. In this case, the mixture model proposed here might be more convenient than the Heligman-Pollard model.

We fit the model also for Italy mortality data since 1872 up to 2008. In order to show the goodness of fit of models we predicted life exepctancy at birth using Siler and mixture model and compare these predictions with real value of e_0. The result is shown in figure 3. It seems that mixture model has a better performance than the Siler model especially in years where accident deaths mode is particularly pronounced. For example in Italy in 1919 the observed value of e_0 was 23.50. The predicted value of Siler model was 35.43 while the predicted value by mixture model was 24.16. However mixture model shows a good fit also when there is no accident hump – in that case the estimate of α is very close to 0, and the SBN distribution approaches to a skew-normal approach. These preliminary results suggest that the proposed model can be a unifying model for all mortality pattern. Siler model cannot catch the accident hump, when it exists – and this might be crucial in many cases, while Heligman-Pollard model is very difficult to fit, especially when accident hump is missing. Moreover all the parameters of mixture model can have a sensible demographic interpretation and in the next future we will try to relate all the parameters to proper life-tables quantities. In addition, We plan to test this model in presence of fragmentary data – a common problem especially for developping countries (see Wheldon et al, [9]), using a Bayesian approach. Moreover, we will use this model to forecast future mortality patterns.

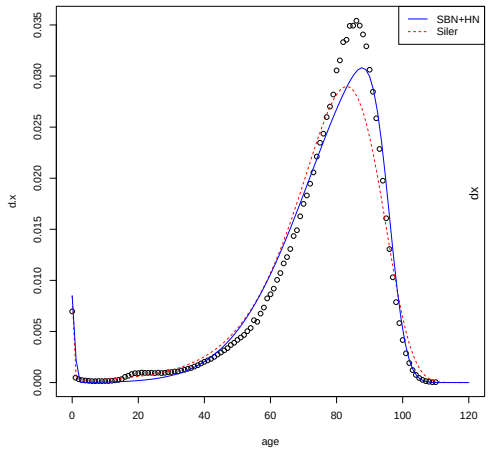

Fig. 1 Men deaths by age (dots) and model fit by Siler model and mixture model. Data :Usa 2002. Source: Human Mortality Database.

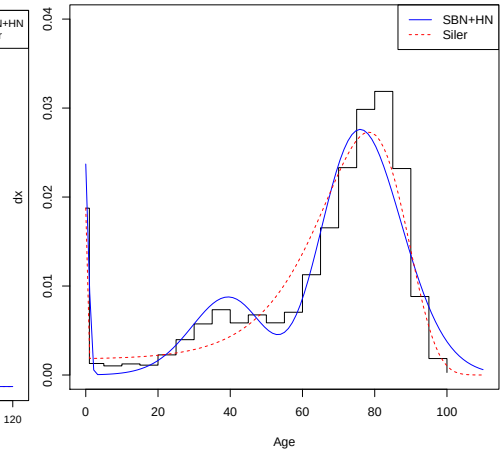

Fig. 2 Women deaths by age (stairs) and model fit by Siler model and mixture model. Data :Botswana 2011. Source: Who Mortality Database.

References

[1] Siu Lan Karen Cheung, Jean-Marie Robin, Edward Jow-Ching Tu, and Graziella Caselli. Three Dimensions of the Survival Curve: Horizontalization, Verticalization, and Longevity Extensions. *Demography*, 42:243–258, 2005.

[2] D. Elal-Olivero, H.W. Gmez, and F.A. Quintana. Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, 139(4):1484–1492, 2009.

[3] L. Heligman and J. Pollard. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1):49–80, 1980.

[4] Gustavo H.M.A. Rocha, Rosangela H. Loschi, and Reinaldo B. Arellano-Valle. Inference in flexible families of distributions with normal kernel. *Statistics: A Journal of Theoretical and Applied Statistics*, 47(6):1184–1206, 2013.

[5] David J. Sharrow, Samuel J. Clark, Mark A. Collinson, Kathleen Kahn, and Stephen M. Tollman. The age pattern of increases in mortality affected by hiv: Bayesian fit of the heligman-pollard model to data from the agincourt hdss field site in rural northeast south africa. *Demographic Research*, 29:1039–1096, December 2013.

[6] William Siler. A competing-risk model for animal mortality. *Ecology*, 60(4):750–757, 1979.

[7] William Siler. Parameters of mortality in human populations with widely varying life spans. *Statistics in Medicine*, 2:373–380, 1983.

[8] University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). *Human Mortality Database*. Available at www.mortality.org or www.humanmortality.de (data downloaded on November 28, 2013).

[9] Mark C. Wheldon, Adrian E. Raftery, Samuel J. Clark, and Patrick Gerland. Estimating demographic parameters with uncertainty from fragmentary data. *Journal of the American Statistical Association*, 108:795–822, 2013.

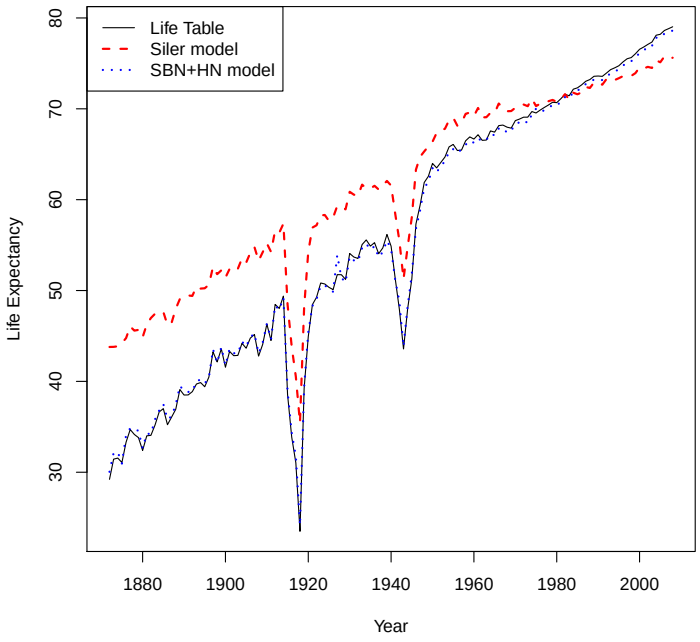

Fig. 3 Life expectancy at birth: Real value and predictions of Siler and mixture model. Data: Italy 1872-2008. Source: Human Mortality Database.