

Introduction

This note describes a methodology for use in graduating population mortality rates for an individual calendar year. It is based on existing techniques in peer-reviewed journals and papers, and is free to use. In addition to smoothing the mortality rates, it is also capable of extrapolating rates to higher ages where data is sparse or even non-existent. A free implementation of this method is available at www.longevity.co.uk/graduate, along with an updated version of this note.

Method

We assume we have a vector of death counts, d_x , where x is the age last birthday. We further assume that we have corresponding central exposed-to-risk data, $e_{x+\frac{1}{2}}$, i.e. mid-age population estimates. We assume that the number of deaths is a random variable, D_x , with a Poisson distribution, i.e.

$$D_x \sim \text{Poisson} \left(e_{x+\frac{1}{2}} \times \mu_{x+\frac{1}{2}} \right)$$

where $\mu_{x+\frac{1}{2}}$ is the force of mortality (hazard rate) applying at age $x + \frac{1}{2}$. For flexibility we use a basis of m B -splines[†] for the force of mortality as follows:

$$\log \mu_{x+\frac{1}{2}} = \sum_{j=1}^m \theta_j B_j \left(x + \frac{1}{2} \right) \quad (1)$$

where $B_j \left(x + \frac{1}{2} \right)$ is the j^{th} B -spline evaluated at $x + \frac{1}{2}$ and the θ_j are coefficients to be estimated. Richards, Kirkby and Currie (2006) give a worked example of how the heights of the B -splines are varied by the θ_j and how the products of the B_j and θ_j are summed to form the value of $\log \mu$. By working on a logarithmic scale, the θ_j are free to vary across the real line.

To estimate the θ_j we form a simple likelihood function for maximisation:

$$L \propto \prod_x \frac{1}{d_x!} \left(e_{x+\frac{1}{2}} \times \mu_{x+\frac{1}{2}} \right)^{d_x} \exp \left(-e_{x+\frac{1}{2}} \times \mu_{x+\frac{1}{2}} \right) \quad (2)$$

although in practice we would maximise the log-likelihood function, l , defined as follows after dropping additive constants involving data only:

$$\ell = \sum_x d_x \log \mu_{x+\frac{1}{2}} - \sum_x e_{x+\frac{1}{2}} \times \mu_{x+\frac{1}{2}} \quad (3)$$

If the spacing between the number of splines is too small we could have an erratic pattern of the θ_j , so we will adapt Equation 3 to include a penalty function as follows:

$$\ell^p = \ell - \lambda P(\theta) \quad (4)$$

where $P(\theta)$ is a *penalty function* to penalise roughness in the θ_j and λ is a parameter controlling the degree of smoothing applied. A common example is to use a second-order penalty function, such the following:

$$P(\theta) = (\theta_1 - 2\theta_2 + \theta_3)^2 + \dots + (\theta_{m-2} - 2\theta_{m-1} + \theta_m)^2 \quad (5)$$

[†] A B -spline is composed piecewise from cubic polynomials, and takes the value of zero outwith a certain range. The polynomials are constructed such that they are smooth at the join points or knots. A B -spline basis is constructed such that these join points occur at the central knot points of other splines, thus making the B -splines overlap with each other. In a cubic B -spline basis, any given point will be covered by at most four B splines — see Richards, Kirkby and Currie (2006) for a detailed discussion and example. The flexibility of a B -spline basis comes from the fact that the height of each B -spline is free to vary independently of all other B -splines.

The expression ℓ^p is known as a *penalized log-likelihood* and, for a given value of λ , the maximum penalized log-likelihood estimate of θ is given by maximising Equation 4. The value of λ can either be pre-set, or else selected by picking the value of λ which minimises an information criterion, such as the AIC, BIC or GCV.

Why structure things this way?

The model is structured around $\log \mu_x$ due to the greater theoretical robustness this gives. By working on a logarithmic scale, we ensure that μ_x is always positive. By working with μ_x , we ensure that any derived q_x will automatically lie in the interval $(0, 1)$ using the standard result:

$$q_x = 1 - \exp \left(- \int_0^1 \mu_{x+s} ds \right) \quad (6)$$

In practice we would use the following approximation:

$$q_x \approx 1 - \exp \left(- \mu_{x+\frac{1}{2}} \right) \quad (7)$$

but the same point holds true: by using μ_x we guarantee that any derived values for q_x lie in $(0, 1)$ without imposing additional checks or constraints. Equally, we can derive any survival probability using the standard result:

$${}_t p_x = \exp \left(- \int_0^t \mu_{x+s} ds \right) \quad (8)$$

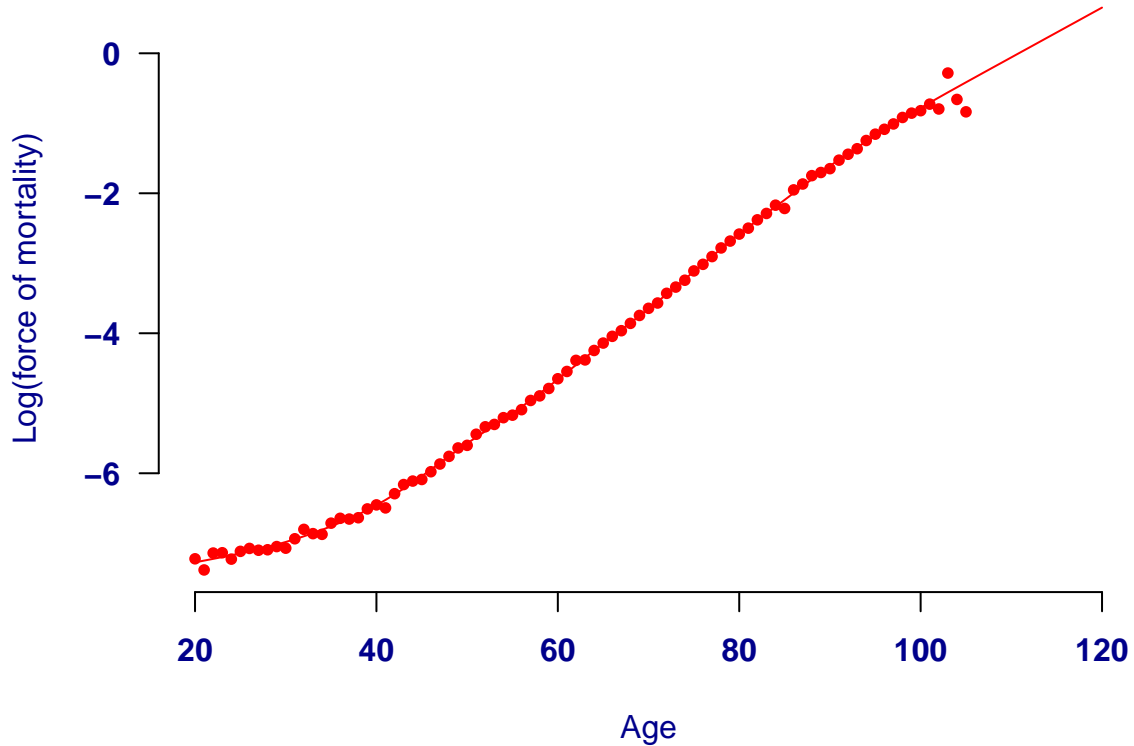
but again in practice we would use the following simple approximation for integer t :

$${}_t p_x \approx \exp \left(- \sum_{i=0}^{t-1} \mu_{x+i+\frac{1}{2}} \right) \quad (9)$$

Some data sets come tailor-made for modelling μ_x : the ONS population data for England and Wales, for example, has death counts at age last birthday in each calendar year, together with mid-year population estimates. Other data sets will have initial population estimates more immediately suitable for a Binomial model of death counts, but the benefits of working with μ_x are such that the exposure data in such instances should be modified to allow modelling of μ_x . This is done by the simple approximation of central exposure by deducting half of the deaths from the initial exposure.

The penalty function in Expression (5) shows a quadratic penalty function, but there are other penalty functions which could potentially be used. However, a quadratic penalty leads to a linear extrapolation of hazard rates on a logarithmic scale, which seems the most appropriate assumption. Figure 1 shows an example of this — the quadratic penalty function produces a linear extrapolation of $\log(\text{mortality hazard})$ to higher ages. Furthermore, a linear function for $\log \mu_x$ means a slower-than-linear increase in $\log q_x$ at older ages — see Richards (2008). This is a desirable feature of extrapolated rates, as such so-called *late-life mortality deceleration* is a commonly observed feature of many populations (Gavrilov and Gavrilova, 2001).

Figure 1. Crude mortality rates for males in England and Wales in 2004 (•) with fitted smooth rates and extrapolation to age 120 (—).



Note that in order to provide a linear extrapolation on a logarithmic scale, it is necessary that the knots are evenly spaced. A knot spacing of five years would be fine for most instances (as in Figure 1), but greater robustness could be ensured by using a spacing of ten years. This sacrifices a small amount of local flexibility in order to gain robustness to wild data values at the uppermost ages.

Advantages of this method

There are several advantages to this method:

1. It is simple. The method can be programmed in a few lines of R, or (with more work) as an Excel spreadsheet.
2. It is flexible — a basis of B -splines can accommodate any reasonable pattern of mortality rates by age.
3. It provides an automatically smooth fit, thus eliminating any “kinks” in the data.
4. It is a statistical method in that the ages with greater numbers of deaths and exposures have the greatest influence on the graduation.
5. As a statistical method it is capable of extrapolating mortality rates by age — see Richards and Currie (2011). This is useful in extending limited data sets to advanced ages, thus enabling the calculation of complete life expectancies instead of temporary life expectancies.
6. Both graduation and extrapolation happen in a single step, rather than two separate processes. Extrapolated rates are therefore consistent with the graduated ones.
7. It is structured around μ_x on a logarithmic scale, which means all derived values for μ_x , q_x and ${}_tp_x$ will automatically be on valid scales.
8. Using a second-order penalty gives a log-linear extrapolation for μ_x . This means a decelerating increase in $\log q_x$, thus giving the late-life mortality deceleration expected of mortality rates at the highest extrapolated ages (Gavrilov and Gavrilova, 2001).

Items to watch

There are some potential pitfalls to avoid:

A. With closely spaced knots it is sometimes possible for extrapolation to exhibit excessive sensitivity to edge effects. This can be avoided by using a wide knot spacing between splines, e.g. 10 years.

B. Over-dispersion. In larger populations the Poisson assumption of variance equalling the mean is often violated and can lead to under-smoothing from small values of λ . This can be allowed for either (i) by including an explicit over-dispersion parameter as in Djeundje and Currie (2011), or (ii) by keeping a wide knot spacing, or (iii) by enforcing a large value of λ .

References

- DJEUNDJE, V. A. B. AND CURRIE, I. D. (2011) Smoothing dispersed counts with applications to mortality data, *Annals of Actuarial Science*, **5**, 33–52.
- GAVRILOV, L. A. AND GAVRILOVA, N. S. (2001) The reliability theory of ageing and longevity, *Journal of Theoretical Biology*, **213**, 513–585.
- RICHARDS, S. J., KIRKBY, J. G. AND CURRIE, I. D. (2006). [The Importance of Year of Birth in Two-Dimensional Mortality Data](#), *British Actuarial Journal*, **12 (I)**, 5–61.
- RICHARDS, S. J. (2008). [Applying survival models to pensioner mortality data](#), *British Actuarial Journal*, **14 (II)**, No. 61, 257–303.
- RICHARDS, S. J. AND CURRIE, I. D. (2011), Extrapolating mortality projections by age, *Life and Pensions Risk magazine*.