BETA KERNEL GRADUATION OF MORTALITY DATA IN R. AN APPLICATION TO THE ENNA PROVINCE

Anna Maria Altavilla, Angelo Mazza, Antonio Punzo

1. Introduction

Mortality rates are age-specific indicators, commonly used in demography. They are also widely adopted by actuaries, in the form of mortality tables, to calculate life insurance premiums, annuities, reserves, and so on. Producing these tables from a suitable set of crude (or raw) mortality rates is called graduation, and this subject has been extensively discussed in the actuarial literature (see, e.g., Copas and Haberman, 1983).

In detail, the crude rates $\dot{q}_x$, for each age x, can be intended as arising from a sample of deaths, of size d_x, from a population, initially exposed to the risk of death, of size e_x, and thus they contain random fluctuations. The relation between and the true - but unknown - mortality rates q_x may be summarized as

$$q_x = \dot{q}_x + \varepsilon_x \ ,$$

where ε_x denotes noise due to random variation, age misstatement, etc.. Such random fluctuations are usually more evident when small area datasets are investigated or at ages in which mortality is itself rarer.

In order to capture the underlying mortality pattern q_x from the crude rates, a graduation function

$$\hat{q}_x = m(\dot{q}_x)$$

is used; $m(\cdot)$ aims at smoothing out irregularities ε_x in crude mortality rates $\dot{q}_x$.

In analogy with the usual statistical modeling, the $m(\cdot)$ function can be specified parametrically (see, *e.g.* Debòn *et al.*, 2005) or nonparametrically (see, *e.g.* Debòn *et al.*, 2006). In this paper, we will present an R computer program for nonparametric graduation using the discrete beta kernel estimator proposed by Mazza and Punzo (2011).

Kernel smoothing is one of the most popular statistical methods for nonparametric graduation. Among the various alternatives existing in literature, the attention is here focused on the discrete beta kernel estimator proposed by Mazza and Punzo (2011). Roughly speaking, the genesis of this model starts with the consideration that, although age X is in principle a continuous variable, it is typically truncated in some way, such as age at last birthday, so that it takes values on the discrete set $\mathcal{X} = \{0,1,...,\omega\}$, ω being the highest age of interest. Note that the discretization of age, from a pragmatic and practical point of view, could also come handy to actuaries that have to produce "discrete" graduated mortality tables starting from the observed counterparts.

In the estimator proposed in Mazza and Punzo (2011), discrete beta distributions are considered as kernel functions, in order to overcome the problem of boundary bias commonly arising from the use of symmetric kernels. The support $\mathcal{X}$ of the discrete beta, which can be asymmetric, in fact matches the age range and this, when smoothing is made near the boundaries, allows avoiding the allocation of weight outside the support (e.g. negative or unrealistically high ages).

An application to mortality data relative to the female population of the city of Enna (Italy) in the year 2009 is presented.

2. The discrete beta kernel estimator

Given the crude rates $\dot{q}_y$, y $\in \mathcal{X}$, the Nadaraya-Watson kernel estimator of the true but unknown mortality rates q_x the evaluation age x is

$$\hat{q}_x = \sum_{y\in\mathcal{X}} \frac{k_h(y;m=x)}{\sum_{j\in\mathcal{X}} k_h(j;m=x)}\dot{q}_y = \sum_{y\in\mathcal{X}} K_h(y;m=x)\dot{q}_y, \quad x \in \mathcal{X}, \tag{1}$$

where $k_h(\cdot;m)$ is the discrete kernel function (hereafter simply named kernel), $m \in \mathcal{X}$ is the single mode of the kernel, $h > 0$ is the so-called bandwidth governing the bias-variance trade-off, and $K_h(\cdot;m)$ is the normalized kernel.

Since we are treating age as being discrete, with equally spaced values, kernel graduation by means of (1) is equivalent to moving (or local) weighted average graduation (Gavin et al., 1995). As kernels, in (1) we use

$$k_h(x;m) = \left(x+\frac{1}{2}\right)^{\frac{m+\frac{1}{2}}{h(\omega+1)}}\left(\omega+\frac{1}{2}-x\right)^{\frac{\omega+\frac{1}{2}-m}{h(\omega+1)}}. \tag{2}$$

The normalized version, $K_h(\cdot;m)$, corresponds to the discrete beta probability mass functions of Punzo and Zini (2012), parameterized according to Punzo (2010, see also Bagnato and Punzo in press), according to the mode m and another parameter h that is closely related to the distribution variability. Substituting (2) in

(1) we obtain the discrete beta kernel estimator that was introduced in Mazza and Punzo (2011).

Roughly speaking, discrete beta kernels possess two peculiar characteristics. Firstly, their shape, fixed h, automatically changes according to the value of m. Secondly, the support of the kernels matches the age range $\mathcal{X}$, so that no weight is assigned outside the data support; this means that the order of magnitude of the bias does not increase near the boundaries. Further details are reported in Mazza and Punzo (2011).

In (1), the parameter h has to be chosen. A low value of h puts more emphasis on fit than smoothness. For values of h close to 0^+, $k_h(x;m)$ tends to a Dirac delta function in $x = m$; that is, there is essentially no graduation, it returns the raw rates. A high value of h puts more emphasis on smoothness than fit: for extremely high values of h, $k_h(x;m)$ tends to a discrete uniform distribution.

The amount of smoothing that is appropriate maybe decided by studying the resulting graduations, in a rather subjective way. Otherwise, ordinary cross-validation may be used; it provides an objective, data-dependent technique for choosing the bandwidth and, although it requires a rather high computational effort, it is, nevertheless, simple to understand and natural for nonparametric regression. In detail, cross-validation simultaneously fits and smooths the data, removing one “data point” at a time, estimating its value and then comparing the estimate to the omitted, observed value. The usual cross-validation statistic or score, $CV(h)$, is

$$CV(h|s) = \sum_{x\in\mathcal{X}} \left(\dot{q}_x - \hat{q}_x^{(-x)}\right)^2 \qquad (3)$$

where

$$\hat{q}_x^{(-x)} = \sum_{\substack{y\in\mathcal{X}\\ y\neq x}} \frac{K_{h_x}(y;m=x)}{\sum_{\substack{j\in\mathcal{X}\\ j\neq x}} K_{h_x}(j;m=x)} \dot{q}_y \ . \qquad (4)$$

is the estimated value at age x computed by removing $\dot{q}_x$.

However, instead of the standard residual sum of squares in (4), Mazza and Punzo (2011) suggest the use of the sum of the squares of the proportional differences

$$CV(h|s) = \sum_{x\in\mathcal{X}} \left(\frac{\dot{q}_x}{\hat{q}_x^{(-x)}} - 1\right)^2 ; \qquad (5)$$

this is a commonly used divergence measure in the graduation literature because, since the high differences in mortality rates among ages, we want the mean relative square error to be low (see Heligman and Pollard, 1980).

Before the model is applied, it may be worth to consider transforming the data into a more tractable form, that more clearly reveals the structure of the data. A commonly used transformation in binary analysis is the logit (or log-odds) transformation

$$\dot{q}_i^t = \ln \frac{\dot{q}_i}{\left(1-\dot{q}_i\right)}. \tag{6}$$

By smoothing on a logistic scale and then back-transforming, we are guaranteed that $0 \leq \hat{q}_x \leq 1$. This transformation also reflects the fact that small changes when the mortality rate is near zero are as important as larger changes when the mortality rate is much higher.

3. Discrete beta kernel graduation using the R statistical environment

This section introduces the essential elements needed for doing the discrete beta kernel graduation using the functions in appendix that we developed for the **R** statistical environment (R Development Core Team, 2011). The main function, **dbkGrad** does the beta kernel graduation. It requires only one argument, the observed mortality rates **obsqx**, while the other arguments are optional. In particular, the **obsqx** argument must be a numeric vector of size $\omega + 1$.

Argument **h** is a scalar; it provides a value for the smoothing parameter. If it is omitted, then the smoothing parameter is computed by means of cross-validation.

Argument **pss** is a logical; if it is **TRUE** then cross-validation selects h by minimizing the proportional sum of squares in (5), while if it is **FALSE** then the sum of square residuals in (3) is minimized; the default value is **TRUE.**

Argument **logit** is a logical; if it is **TRUE** then the logit transformation in (6) is applied to the data before graduating, and then data are back-transformed to obtain the estimate of the true rates; its default value is **FALSE**.

Finally, argument **omega** is a scalar; it sets the upper age limit and its default value is the length of the vector **obsqx** minus one.

In the cross-validation routine, minimization is performed using the Levenberg-Marquardt nonlinear least-squares algorithm, as implemented in the package **minpack.lm** (Elzhov et al., 2010); this package has to be installed before running our code.

4. An application to the Enna province

The **dbkGrad** function is applied to mortality data relative to the female population of the city of Enna (Italy) in the year 2009, up to 85 years old (source: Istat, 2012).

To begin the analysis, the program and data have to be loaded; if we assume that within the **R** working directory there are a text file named “dbkGrad.R” containing the code in appendix and a table format file named “enna2009female” containing the raw dataset, this may be done with the commands

```
R> source("dbkGrad.R")
R> rawdata <- as.vector(read.table("enna2009female"))
```

The command used to perform the graduation is:

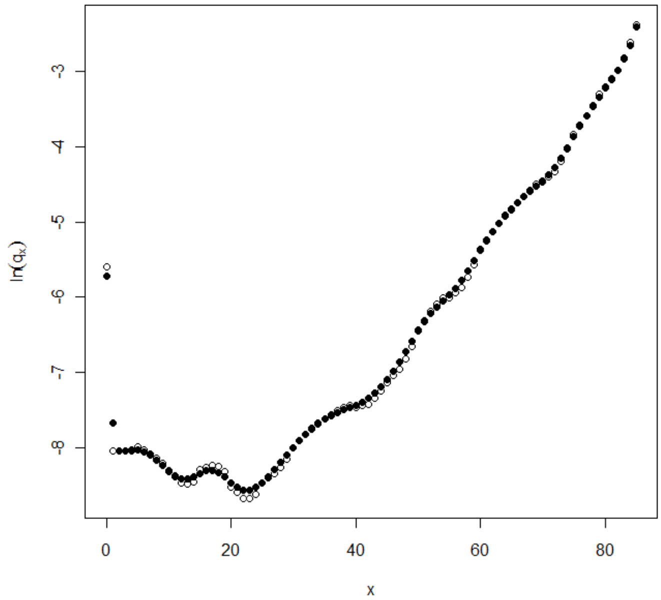
```
R> grad_data <- dbkGrad(obsqx=rawdata)
```

Here, the default values have been used for all the omitted arguments, so cross-validation has been used to select the smoothing parameter minimizing the proportional sum of squares (**pss=TRUE**), **omega** has been set at the highest observed age (**length(obsqx)-1),** logit transformation has not been applied (**logit=FALSE**).

Both the observed and graduated data are depicted in Fig. 2 using a logarithmic scale. Since $\dot{q}_x$ from age 25 to about 80 grows at an approximately exponential rate, with the logarithmic scale data points in that region almost lay over a straight line. Plots were obtained using the standard plot function, as:

```
R> omega <- length(rawdata)-1
R> plot(0:omega, log(grad_data), lwd=1,
   + type="p", ylab=expression(ln(q[x])))
R> points(0:omega,log(rawdata), lty=5, lwd=1, pch=19)
```

Figure 1 – *Observed (◦) and graduated (•) female mortality rates, in logarithmic scale for the year 2009 in the Enna province.*



5. Conclusions

In this paper we have discussed an **R** function specifically conceived for nonparametric graduation of discrete finite functions, such are age-dependent mortality data, proposed by Mazza and Punzo (2011). Over other graduation techniques, this one has the advantage that kernel functions are chosen from a family of conveniently discretized and re-parameterized beta densities; since their support matches the age range boundaries, the estimates are free of boundary bias.

The use of the proposed **R** function is conceptually simple, and all of its arguments are optional except, obviously, for the raw data vector.

An application to 2009 mortality data for the province of Enna (Italy) has been proposed. Since the area considered is relatively small, random variations made raw data slightly noisy. Graduated data, on the other side, were smooth and consistent.

Appendix

The following code implements the Discrete Beta Kernel Graduation in the **R** statistical environment (R Development Core Team, 2011).

```
dbkGrad<-function(obsqx, pss=T, logit=F, h=NULL, omega=NULL){
    if(is.null(omega)) omega <- length(obsqx)-1
    obsqx <- obsqx[0:(omega+1)]
    if(logit) obsqx <- log(obsqx/(1-obsqx))
    if(is.null(h)){
      library("minpack.lm")
      h <- nls.lm(par=0.01, pss=pss, omega=omega, obsqx=obsqx,
         fn=dbkCV, control = nls.lm.control(maxiter=1000))$par
    }
    K <- dbkern(h,omega)
    Kernels <- K/rowSums(K)
    qxest <- Kernels %*% obsqx
    if(logit) qxest <- exp(qxest)/(1+exp(qxest))
    return(list(qxest=qxest,h=h))
}
dbkern<-function(bandwidth,omega) {
  x  <- m <- 0:omega
  K  <- array(0,c(omega+1,omega+1),dimnames=list(x,m))
  f  <- function(i,j){
    y <- (x[j]+0.5)/(omega+1)
    a <- (m[i]+0.5)/(bandwidth*(omega+1))+1
    b <- (omega+0.5-m[i])/(bandwidth*(omega+1))+1
   dbeta(x=y, shape1=a, shape2=b, ncp=0, log=F)
  }
  gridi <- gridj <- 1:(omega+1)
  K <- outer(gridi,gridj,f)
  return(K)
}
dbkCV <- function(par, bandwidth, omega, obsqx, pss){
    K <- dbkern(par,omega)
    Kd <- K - diag(diag(K))
    Kdrop <- Kd/rowSums(Kd)
    obsqxremove <- Kdrop %*% obsqx
    if(pss) return (obsqxremove/obsqx-1)
    else return (obsqxremove-obsqx)
}
```

References

BAGNATO L, PUNZO A (in press). Finite Mixtures of Unimodal Beta and Gamma Densities and the *k*-Bumps Algorithm. *Computational Statistics*.

COPAS J. B., HABERMAN, S. 1983. Non-parametric graduation using kernel methods. *Journal of the Institute of Actuaries*, 110, pp.135–156.

DEBÒN A., MONTES, F., SALA, R. 2006. A comparison of nonparametric methods in the graduation of mortality: Application to data from the Valencia region (Spain). *International Statistical Review*, 74(2), pp.215–233.

ELZHOV TV, MULLEN KM, BOLKER B 2010. minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK. *R package version 1.1-6*,

GAVIN, J., HABERMAN S., AND VERRALL R. 1995. Graduation by kernel and adaptive kernel methods with a boundary correction. *Transactions of the Society of Actuaries*, 47, pp.173–209.

HELIGMAN L., POLLARD J. 1980. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107, pp.49–80.

ISTAT 2012. Tavole di Mortalità della popolazione italiana per provincia e regione di residenza. http://demo.istat.it/

MAZZA A., PUNZO A. 2011. Discrete beta kernel graduation of age-specific demographic indicators. In INGRASSIA S., ROCCI R., VICHI M., (Eds), *New Perspectives in Statistical Modeling and Data Analysis*, Springer, pp. 127–134.

PUNZO A. 2010. Discrete Beta-type Models. In LOCAREK-JUNGE H, WEIHS C (eds.), *Classification as a Tool for Research*, Springer, pp. 253–261.

PUNZO A., ZINI A (2012). Discrete Approximations of Continuous and Mixed Measures on a Compact Interval. *Statistical Papers*, 53(3), 563-575.

R Development Core Team 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna: Austria.

SUMMARY

Beta kernel graduation of mortality data in R. An application to the Enna province

Various approaches have been proposed in literature for the kernel graduation of mortality rates. Among them, this paper considers the fixed bandwidth discrete beta kernel estimator, a recent proposal conceived to intrinsically reduce boundary bias and in which age is pragmatically considered as a discrete variable. In this paper, we present an implementation of this estimator for the R statistical environment. An application to 2009 female mortality data from Enna (Italy) is also presented.

Anna Maria ALTAVILLA, Professor, University of Catania, Department "Economia ed Impresa", altavil@unict.it

Angelo MAZZA, Assistant Professor, University of Catania, Department "Economia ed Impresa", a.mazza@unict.it

Antonio PUNZO, Assistant Professor, University of Catania, Department "Economia ed Impresa", antonio.punzo@unict.it